AN INVESTIGATION OF THE LINEAR ELECTRO-OPTIC EFFECT OF EPITAXIAL ZINC SULFIDE FILMS FOR INTEGRATED OPTICS

By

JOHN FRANKLIN EBERSOLE

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

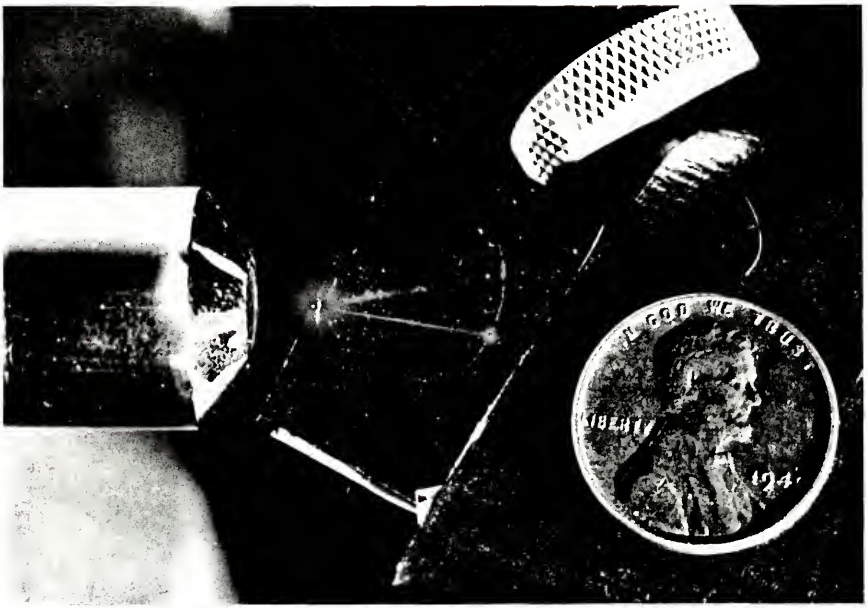
1974

Copyright

By

John Franklin Ebersole

1974



This is a triple-exposure photograph showing a 10X microscope objective (to the left), a gallium arsenide rectangular substrate with a semi-circular film of zinc sulfide, and a penny for size comparison. Also shown is a red (λ = 6328 Å) laser beam in two positions. The upper beam (in air) is visible due to scatter from surface dust particles on the ZnS-GaAs sample as well as from smoke used to help indicate the path of the beam. This beam (slightly off-axis with respect to the 10X objective axis) represents light rays undeviated by the film. The lower beam represents light refracted away from the undeviated beam according to Snell's Law into the zinc sulfide when the film is placed at the focal point of the objective. The beam is guided by the film and propagates until it reaches the other end of the semi-circle, at which point the light radiates brightly. That this lower beam is visible may be attributed to surface irregularities as well as to internal scattering from ZnS crystal imperfections.

To my wife, Ingrid

ACKNOWLEDGMENTS

I wish to express my appreciation to Professor Stanley S. Ballard, Supervisory Committee Chairman, for his valuable professional advice and guidance during my years as a graduate student. I am very grateful to Itek Corporation for providing the opportunity, facilities, and equipment for this research. Special appreciation is expressed to Dr. Ralph E. Aldrich, Dr. Richard H. Hudgin, and Dr. James C. Wyant for their many helpful suggestions concerning the various phases of my research and for their editing of this manuscript. An expression of gratitude goes to Dr. Julius Feinleib who suggested the idea for this research project. The continued encouragement and advice of Dr. Robert Hills, Jr., and Dr. F. Dow Smith are gratefully acknowledged. The technical discussions with Dr. Stephen G. Lipson, Mr. Donald W. Wilmot, and Mr. Paul Vohl, the assistance of Mr. James B. McCallum with crystal preparation, the skillful typing of Mrs. Kathleen Gerow, and the excellent reproduction of this manuscript by Mr. Lawrence W. Gay and his staff have all contributed immensely to the success of this research effort. The support of Mr. Richard T. Turpin, Mr. Robert S. Hilbert, Mr. Walter S. Crowell, Mr. William J. Davis, and Mr. Richard J. Wollensak is greatly appreciated. Finally, I wish to acknowledge my former undergraduate optics and physics professor, Dr. Roy C. Gunter, Jr., and my former high school physics teacher, Mr. Harold F. Desmond; their inspiration has in a very large way contributed to the achievement of my scientific and educational goals.

PREFACE

On June 26, 1972 Dr. Stanley S. Ballard, Professor of Physics, University of Florida, visited the Optical Systems Division of Itek Corporation and met with this author and with Dr. Julius Feinleib, Manager of the Physics Scientific Staff (part of Itek's Central Research Laboratories). Dr. Feinleib described Itek's capability to fabricate cubic-structure zinc sulfide films on gallium arsenide substrates. As a result of that meeting, it was agreed that ZnS-on-GaAs samples would be made available for electro-optical waveguide experiments.

Subsequently, preliminary experiments and analyses were performed, and a proposed line of research was submitted to the author's Ph.D. supervisory committee. That proposal suggested both a theoretical and an experimental research effort. The theoretical part was to concentrate on the propagation of light in inhomogeneous waveguides. The experimental part was to examine the electro-optic effect of zinc sulfide waveguides. This dissertation presents the results of research encompassing both aspects of the proposed research.

TABLE OF CONTENTS

Page

FRONTISPIECE iii

ACKNOWLEDGMENTS v

PREFACE vi

ABSTRACT x

I. INTRODUCTION 1

II. BASIC PHYSICS OF HOMOGENEOUS ASYMMETRIC PLANAR OPTICAL WAVEGUIDES 2

Introduction 2

Vector Potentials 2

TE Modes 5

TM Modes 9

TE + TM Modes 11

Quasi-Geometrical Ray-Optics Approach 12

III. BASIC PHYSICS OF INHOMOGENEOUS ASYMMETRIC PLANAR OPTICAL WAVEGUIDES 17

Introduction 17

Quasi-Geometrical Ray-Optics Approach 17

Solutions for $\vec{E}$ and $\vec{H}$ 26

IV. BASIC PHYSICS OF TOTAL INTERNAL REFRACTION IN INHOMOGENEOUS PLANAR OPTICAL WAVEGUIDES 29

Introduction 29

Page

Quasi-Geometrical Ray-Optics Approach..... 29

Solutions for $\vec{E}$ and $\vec{H}$..... 40

1. The Linear Approximation..... 40

2. TE Modes..... 42

3. TM Modes..... 48

V. GUIDING OF LIGHT IN EPITAXIAL ZINC SULFIDE FILMS..... 54

Introduction..... 54

Total Internal Refraction..... 54

Partial Internal Reflection..... 56

VI. THE LINEAR ELECTRO-OPTIC EFFECT OF EPITAXIAL ZINC SULFIDE FILMS 60

The Principal Axes..... 60

The Transverse Electro-Optic Effect..... 63

Intensity Modulation for Propagation in the ($0\bar{1}1$) Direction 65

1. TE Modes..... 65

2. TM Modes..... 66

3. TE + TM Modes..... 66

Intensity Modulation for Propagation in the (001) Direction..... 71

1. TE Modes..... 71

2. TM Modes..... 71

3. TE + TM Modes..... 72

VII. EXPERIMENTAL PROCEDURE..... 74

Sample Preparation..... 74

Experimental Set-up..... 77

Page

VIII. EXPERIMENTAL RESULTS AND OBSERVATIONS..... 81

Light Guiding..... 81

Electro-Optic Modulation..... 92

IX. SUMMARY..... 108

APPENDIX A. APPLICATION OF MAXWELL'S EQUATIONS AND VECTOR POTENTIALS TO HOMOGENEOUS AND INHOMOGENEOUS PLANAR OPTICAL WAVEGUIDES..... 110

Maxwell's Equations..... 110

Vector Potentials..... 113

Homogeneous Waveguides..... 116

Inhomogeneous Waveguides..... 122

APPENDIX B. HOMOGENEOUS ASYMMETRIC WAVEGUIDES..... 127

TE Modes..... 127

TM Modes..... 143

APPENDIX C. INHOMOGENEOUS ASYMMETRIC WAVEGUIDES..... 157

TE Modes..... 157

TM Modes..... 179

APPENDIX D. TOTAL INTERNAL REFRACTION IN INHOMOGENEOUS WAVEGUIDES..... 193

Introduction..... 193

TE Modes..... 195

TM Modes..... 210

APPENDIX E. THE PRINCIPAL AXES AND INDEXES OF CUBIC ZINC SULFIDE..... 220

LIST OF REFERENCES..... 225

BIOGRAPHICAL SKETCH..... 227

Abstract of Dissertation Presented to the Graduate Council
of the Univiversity of Florida in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

AN INVFSTIGATION OF THE LINEAR ELECTRO-OPTIC EFFECT OF FPITAXIAL ZINC SULFIDE FILMS FOR INTEGRATED OPTICS

By

JOHN FRANKLIN EBERSOLE

March, 1974

Chairman: Stanley S. Ballard
Major Department: Physics

The basic physics of homogeneous asymmetric waveguides is reviewed, allowing ready comparison of similar concepts between homogeneous and inhomogeneous optical waveguides. It is shown that a quasi-geometrical ray-optics approach leads to the same results obtained from more rigorous mathematics, permitting a practical understanding of the phenomenon of total internal refraction and/or reflection in general-index-gradient planar optical waveguides. The phase change of $\pi/2$ (not 0 or π) occurring upon total internal refraction is explained as being due to a cylindrical focussing of the light within the guide. Waveguiding and electro-optic modulation of light in epitaxial zinc sulfide films on gallium arsenide substrates are also discussed. Two mechanisms for light guiding at the ZnS/GaAs boundary are described: total internal refraction or partial internal reflection. From the discovery of a very thin, transparent, low-index inhomogeneous layer of zinc sulfide near the boundary, it is concluded that 1) 90-95% of the zinc sulfide thin film is homogeneous, and 2) due to this thin inhomogenous laver, the zinc sulfide guide is isolated from the gallium arsenide substrate; therefore the guide is not a leaky one but instead is a true dielectric waveguide. Finally, electro-optic measurements are presented which agree with those predicted from theory.

CHAPTER I

INTRODUCTION

This dissertation begins with a discussion of homogeneous waveguides (even though they have been treated by other authors) and develops two systematic means for understanding the phenomenon of total internal refraction and/or reflection in inhomogeneous waveguides; a quasi-geometrical ray-optics approach, and a more rigorous approach using electric and magnetic vector potentials to solve Maxwell's equations. Having done this, the theoretical and physical-insight discussions of the inhomogeneous case should present no problems, for although the mathematics are more complicated and more tedious, the approach is basically the same as in the homogeneous case.

Because of the complexity of the mathematics in the theoretical part, the long derivations have been relegated to several appendices. Then, in the main body, summaries of the important results are presented with physically based explanations, permitting a practical understanding of the propagation of light in general-index-gradient planar optical waveguides.

After homogeneous and inhomogeneous waveguides are discussed, the results of waveguiding and electro-optic light modulation experiments with epitaxial zinc sulfide films are presented.

CHAPTER II

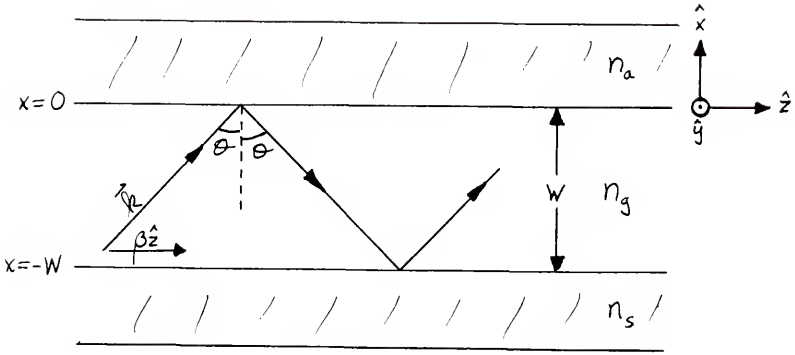
BASIC PHYSICS OF HOMOGENEOUS ASYMMETRIC PLANAR OPTICAL WAVEGUIDES

Introduction

A planar or slab (as opposed to a circular fiber) optical waveguide is one that confines light in one dimension only, as shown in Figure II-1. Light propagating in the $+\hat{z}$ direction is guided or confined by the two surfaces at $x = 0, -W$ by total internal reflection. There is no confinement in the $\hat{y}$ direction. A homogeneous optical waveguide is one that has an index of refraction n_g which is constant throughout the guide. An asymmetric guide has a substrate refractive index n_s which is not equal to the superstrate index n_a, i.e., $n_s \neq n_a$. With no loss of generality, n_s is assumed to be greater than n_a, i.e., $n_s > n_a$. A large number of optical waveguides can be classified as homogeneous asymmetric planar waveguides.

Vector Potentials

In order to describe the propagation of light in optical waveguides, the method of dual vector potentials[1,2] is reviewed in Appendix A. Beginning with Maxwell's equations, the wave equations for $\vec{E}$ and $\vec{H}$ are obtained and the form of solution is shown. Then, for the case where no charge density or currents are present, it is shown that both $\vec{E}$ and $\vec{H}$ can be expressed in terms of a magnetic vector potential $\vec{A}$ and an electric vector potential $\vec{F}$. The wave equations for $\vec{A}$ and $\vec{F}$ for both the homogeneous and inhomogeneous cases are



$$n_g > n_s > n_a$$

$$\beta = k_0 n_g \sin\theta$$

Figure II-1. Cross-sectional view of light propagating in a homogeneous asymmetric planar optical waveguide.

$$\nabla^2\vec{A} - \left(\frac{1}{n^2}\right)\left(\nabla\cdot\vec{A}\right)\left(\nabla n^2\right) + k_o^2 n^2 \vec{A} = 0 \qquad \text{II-(1)}$$

$$\nabla^2\vec{F} + k_o^2 n^2 \vec{F} = 0 \qquad \text{II-(2)}$$

Of course, for the homogeneous case, $\nabla n^2 = 0$ in Eq. (1). Here $k_o = 2\pi/\lambda_o = \omega/c$, λ_o is the free-space wavelength, ω is the angular frequency, and c is the speed of light in free space. The following expressions result for $\vec{E}$ and $\vec{H}$:

$$\vec{E} = -\nabla\times\vec{F} - \left(i/k_o n^2\right)\nabla\left(\nabla\cdot\vec{A}\right) - i k_o \vec{A} \qquad \text{II-(3)}$$

$$\vec{H} = \nabla\times\vec{A} - \left(i/k_o\right)\nabla\left(\nabla\cdot\vec{F}\right) - i k_o n^2 \vec{A} \qquad \text{II-(4)}$$

For the choice of potential

$$\vec{A} = \hat{x}\, A_x \qquad \text{II-(5)}$$

$$\vec{F} = \hat{x}\, F_x \qquad \text{II-(6)}$$

(where $\hat{x}$, $\hat{y}$, and $\hat{z}$ are Cartesian unit normals), it is shown that

$$E_x = \left(i/k_o n^2\right)\frac{\partial^2 A_x}{\partial z^2} \qquad \text{II-(7)}$$

$$E_y = -\frac{\partial F_x}{\partial x} \qquad \text{II-(8)}$$

$$E_z = -\left(i/k_o n^2\right)\frac{\partial^2 A_x}{\partial z\,\partial x} \qquad \text{II-(9)}$$

$$H_x = \left(i/k_o\right)\frac{\partial^2 F_x}{\partial z^2} \qquad \text{II-(10)}$$

$$H_y = \frac{\partial A_x}{\partial x} \qquad \text{II-(11)}$$

$$H_z = -(i/k_0)\frac{\partial^2 F_x}{\partial z \partial x} \qquad \text{II-(12)}$$

For the choice

$$\vec{A} = 0 \qquad \text{II-(13)}$$

$$\vec{F} = \hat{x} F_x \qquad \text{II-(14)}$$

then $E_x = 0$, $F_z = 0$, and $H_y = 0$, in which case, because there is no $\hat{z}$-component of $\vec{E}$, a pure transverse electric (TE) mode of propagation exists. For the choice

$$\vec{A} = \hat{x} A_x \qquad \text{II-(15)}$$

$$\vec{F} = 0 \qquad \text{II-(16)}$$

then $E_y = 0$, $H_x = 0$, $H_z = 0$, and a pure transverse magnetic (TM) mode of propagation exists. It may be seen that a field of any polarization in an asymmetric planar optical waveguide can be expressed as the sum (i.e., the superposition) of a TE field and a TM field.

TE Modes

In Appendix B, homogeneous asymmetric waveguides are discussed in in detail for both TE and TM modes. After an analysis involving boundary conditions, simultaneous equations, and complex algebra, the

following solutions for $\vec{E}$ and $\vec{H}$ are obtained for TE modes:

$$\vec{E}_{TE} = \hat{y}\, E_y \qquad \text{II-(17)}$$

$$\vec{H}_{TE} = \hat{x}\, H_y + \hat{z}\, H_z \qquad \text{II-(18)}$$

$$E_y(x,z,t) = E_{oy}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(19)}$$

$$H_x(x,z,t) = H_{ox}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(20)}$$

$$H_z(x,z,t) = H_{oz}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(21)}$$

Here β is the wavevector component along the $\hat{z}$-axis, i.e.,

$$\beta = k_o n_g \sin\theta = \text{constant} \qquad \text{II-(22)}$$

and has a constant value in all three of the n_a, n_g, and n_s regions. Eq. (22) implies that the wave travels at the same velocity v_z in all three regions since $v_z = c\beta$. Finally,

$$E_{oy}(x) = \begin{cases} 2F_2\, e^{-i\varphi_{ag}^{TE}} \cos\left(\varphi_{ag}^{TE}\right) e^{-ax} & \text{for } x \geqq 0 \qquad \text{II-(23)} \\ 2F_2\, e^{-i\varphi_{ag}^{TE}} \cos\left(gx + \varphi_{ag}^{TE}\right) & \text{for } -W \leqq x \leqq 0 \qquad \text{II-(24)} \\ 2F_2\, e^{-i\varphi_{ag}^{TE}} \cos\left(\varphi_{gs}^{TE} + m\pi\right) e^{s(W+x)} & \text{for } x \leqq -W \qquad \text{II-(25)} \end{cases}$$

and, from Eqs. (8), (10), and (12), $H_x = (-i/k_o)\partial E_y/\partial z$ and $H_z = (i/k_o \partial E_y/\partial x)$. The exact form of H_{ox} and H_{oz} is given in Appendix B, Eqs. B-(83) through B-(86). These values for E_{oy}, H_{ox}, and H_{oz} show that the wave is confined to the n_g region (the waveguide) and that the fields go to zero as $x \to \pm \infty$. The coefficient F_2 is arbitrary, being determined, for instance, by how much light is coupled into the guide. The remaining parameters as determined in Appendix B are given below:

$$a = k_o \left(n_g^2 \sin^2\theta - n_a^2\right)^{1/2} \qquad \text{II-(26)}$$

$$g = k_o n_g \cos\theta \qquad \text{II-(27)}$$

$$s = k_o \left(n_g^2 \sin^2\theta - n_s^2\right)^{1/2} \qquad \text{II-(28)}$$

$$\phi_{ag}^{TE} = \tan^{-1}(a/g) \qquad \text{II-(29)}$$

$$\phi_{gs}^{TE} = \tan^{-1}(g/s) \qquad \text{II-(30)}$$

In the course of obtaining the solutions for $\vec{F}_{TE}$ and $\vec{H}_{TE}$, two useful TE equations were obtained:

$$2W k_o n_g \cos\theta - 2\phi_{ag}^{TE} - 2\phi_{gs}^{TE} = 2m\pi \qquad \text{II-(31)}$$

$$(W_{min})_{TE} = \frac{\lambda_o}{2\pi \left(n_g^2 - n_s^2\right)^{1/2}} \left\{ m\pi + \tan^{-1}\left[\frac{\left(n_s^2 - n_a^2\right)^{1/2}}{\left(n_g^2 - n_s^2\right)^{1/2}}\right] \right\} \qquad \text{II-(32)}$$

where $(W_{min})_{TE}$ is the minimum waveguide thickness required in order to propagate a mode of order m. Eq. (31) is called the characteristic waveguide equation.

Eqs. (31) and (32) show that, depending on λ_o, n_a, n_g, and n_s, only certain modes can propagate in a film of thickness W. Further, for $n_s \neq n_a$, there is a cutoff thickness, W_{min}, below which no guided wave can be propagated, even for m = 0. It may be noted that Eq. (31) is a transcendental equation in Θ.

Fq. (32) is derived from Fq. (31) using the following argument. In order for the waves to leak out or escape (i.e., for propagation to cease), the waves must escape either into the n_a - region or into the n_s - region. For such a situation, at least one of the functions e^{-ax} [for $x \geq 0$] or e^{sx} [for $x \leq -W$] in Eqs. (23) and (25) must no longer be decaying exponentials (which are real functions) but must instead be purely imaginary. From Eqs. (26) and (28), it may be seen that when $n_g \sin\Theta < n_a$ or when $n_g \sin\Theta < n_s$, then e^{-ax} or e^{sx} respectively is imaginary. Thus when either

$$\sin\Theta = n_a/n_g \qquad \text{II-(33)}$$

or

$$\sin\Theta = n_s/n_g \qquad \text{II-(34)}$$

then waves will begin to leak out. When $\sin\Theta = n_s/n_g$, then Θ is at the critical angle for total internal reflection at the $n_g - n_s$ boundary. Since $n_s > n_a$, total reflection still occurs at the $n_a - n_g$ boundary. When $\sin\Theta < n_s/n_g$, plane waves escape into the n_s medium,

and propagation ceases, i.e., the mode is cut off. Using Eq. (34) in Eq. (31) results in Eq. (32).

Referring to Eq. (31), it may be seen that Θ can never reach the value of $\pi/2$ (propagation straight down the guide), since in the limit as $\Theta \to \pi/2$, $W \to \infty$, which is physically unrealistic. Put another way, no longer is there really any waveguiding as $\Theta \to \pi/2$. Combining this fact with Eq. (34) results in the relationship

$$n_g > n_g \sin\Theta > n_s \qquad \text{II-(35)}$$

as the range of values Θ for which waveguiding (without cutoff) can occur, as determined by Eq. (31).

TM Modes

Appendix B also contains a derivation of $\vec{E}$ and $\vec{H}$ for TM modes propagating in a homogeneous asymmetric waveguide. The solutions are

$$\vec{E}_{TM} = \hat{x}\, E_x + \hat{z}\, E_z \qquad \text{II-(36)}$$

$$\vec{H}_{TM} = \hat{y}\, H_y \qquad \text{II-(37)}$$

$$E_x(x,z,t) = E_{ox}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(38)}$$

$$E_z(x,z,t) = E_{oz}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(39)}$$

$$H_y(x,z,t) = H_{oy}(x)\, e^{i(\omega t - \beta z)} \qquad \text{II-(40)}$$

Here β has the value given in Eq. (22) and is a constant. The x dependence is given by

$$H_{oy}(x) = \begin{cases} -2A_2 e^{-i\phi_{ag}^{TM}} \cos\left(\phi_{ag}^{TM}\right) e^{-ax} & \text{for } x \geqq 0 \quad \text{II-(41)} \\ -2A_2 e^{-i\phi_{ag}^{TM}} \cos\left(gx + \phi_{ag}^{TM}\right) & \text{for } -W \leqq x \leqq 0 \quad \text{II-(42)} \\ -2A_2 e^{-i\phi_{ag}^{TM}} \cos\left(\phi_{gs}^{TM} + m\pi\right) e^{s(W+x)} & \text{for } x \leqq -W \quad \text{II-(43)} \end{cases}$$

From Eqs. (7), (9), and (11), $E_x = (i/k_o n^2)\partial H_y/\partial z$ and $E_z = (-i/k_o n^2)\partial H_y/\partial x$. The exact form of E_{ox} and E_{oz} is given in Eqs. B-(157) through B-(162). The coefficient A_2 is arbitrary, as was the coefficient F_2 in the TE case. The parameters a, g, and s are given by Eqs. (26) - (28). The quantities ϕ_{ag}^{TM} and ϕ_{gs}^{TM} are given by

$$\phi_{ag}^{TM} = \tan^{-1}\left[\left(n_g^2/n_a^2\right)(a/g)\right] \qquad \text{II-(44)}$$

$$\phi_{gs}^{TM} = \tan^{-1}\left[\left(n_g^2/n_s^2\right)(s/g)\right] \qquad \text{II-(45)}$$

It may be noted from Eqs. (29)-(30) and (44)-(45) that

$$\tan \phi_{ag}^{TM} = \left(n_g^2/n_a^2\right) \tan \phi_{ag}^{TE} \qquad \text{II-(46)}$$

$$\tan \phi_{gs}^{TM} = \left(n_g^2/n_s^2\right) \tan \phi_{gs}^{TE} \qquad \text{II-(47)}$$

In the course of obtaining the solutions for $\vec{E}_{TM}$ and $\vec{H}_{TM}$ in Appendix B, the characteristic waveguide and cutoff equations for the

TM case were obtained:

$$2Wk_o n_g \cos\theta - 2\phi_{ag}^{TM} - 2\phi_{gs}^{TM} = 2m\pi \qquad \text{II-(48)}$$

$$(W_{min})_{TM} = \frac{\lambda_o}{2\pi (n_g^2 - n_s^2)^{1/2}} \left\{ m\pi + \tan^{-1} \left[\frac{n_g^2}{n_a^2} \frac{(n_s^2 - n_a^2)^{1/2}}{(n_g^2 - n_s^2)^{1/2}} \right] \right\} \qquad \text{II-(49)}$$

Eq. (49) is similar to Eq. (32) except for the factor n_g^2/n_a^2 in the argument of the arctangent. From Eqs. (48) and (49), it may be seen that the value of W for the TM case must be larger than W in the TE case from Eqs. (31) and (32), owing to the fact that $\phi_{ag}^{TM} > \phi_{ag}^{TE}$ and $\phi_{gs}^{TM} > \phi_{gs}^{TE}$. Finally, Eq. (35) applies equally well to TM modes for the allowable values of θ for waveguiding without cutoff, as determined by Eq. (48).

TE + TM Modes

As mentioned above and shown in more detail in Appendix A, a field of any polarization in a homogeneous asymmetric planar guide can be expressed as the sum (i.e., the superposition) of a TE field and a TM field. Further, the components obtained from the TE derivation [E_y, H_x, and H_z] obey the waveguide and cutoff equations given by Eqs. (31) and (32), whereas the components obtained from the TM derivation [E_x, E_z, and H_y] obey Eqs. (48) and (49). For propagation of a TE + TM wave in a homogeneous guide of thickness W, the angle $\theta = \theta_{TE}$ in Eq. (31) is not equal to the angle $\theta = \theta_{TM}$ in Eq. (48) due to the different phases ϕ_{ag}^{TE}, ϕ_{ag}^{TM}, ϕ_{gs}^{TE}, and ϕ_{gs}^{TM}. As a consequence, the propagation constant β must be different for the TE and TM cases, i.e.,

$$\beta_{TE} = k_o n_g \sin\theta_{TE} \qquad \text{II-(50)}$$

$$\beta_{TM} = k_o n_g \sin \theta_{TM} \qquad \text{II-(51)}$$

Hence, as the TE + TM wave propagates down the guide according to $e^{-i\beta z}$, the phase difference $\Delta\beta$ between the TE and the TM parts of the TE + TM wave is

$$\Delta\beta = (\beta_{TE} - \beta_{TM})\, L \qquad \text{II-(52)}$$

$$= k_o n_g (\sin\theta_{TE} - \sin\theta_{TM})\, L \qquad \text{II-(53)}$$

where L is the length of the guide. Thus a TE + TM wave, initially linearly polarized, will become more or less elliptically polarized, depending on θ_{TE} and θ_{TM} [which are determined by Eqs. (31) and (48)] and L, for a given k_o and n_g.

Concerning the minimum thickness $(W_{min})_{TE + TM}$ necessary to propagate a given mode, the larger of $(W_{min})_{TE}$ or $(W_{min})_{TM}$ [given by Eqs. (32) and (49)] must be the minimum thickness. Owing to the factor n_g^2/n_a^2 in the argument of the arctangent of Eq. (49), $(W_{min})_{TM}$ is the larger (since $n_g > n_a$). Hence the cutoff thickness for TE + TM modes is

$$(W_{min})_{TE+TM} = (W_{min})_{TM} \qquad \text{II-(54)}$$

Quasi-Geometrical Ray-Optics Approach

It is possible to derive the waveguide equations using a quasi-geometrical approach similar to one given by Tien and Ulrich.[3] Referring to Figure II-2, the plane wave propagating down the guide may be thought of as a zig-zagging within the guide, plus an evanescent wave

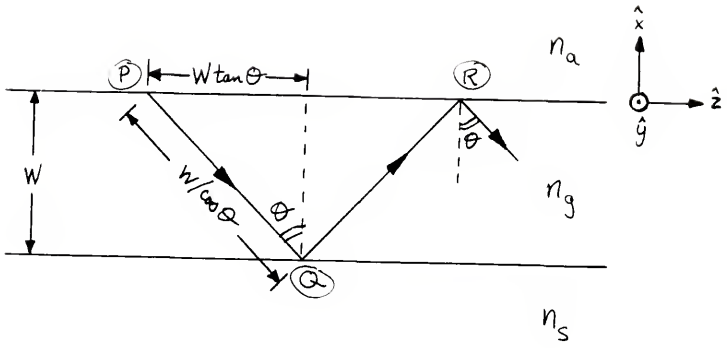


Figure II-2. Geometry of a homogeneous optical waveguide.

propagating in the superstrate and in the substrate at the same velocity in the $\hat{z}$-direction as in the guide. This is consistent with the statement made earlier in this chapter that the wavevector component along the $\hat{z}$-axis, β, has a constant value $k_o n_g \sin\Theta$ in all three regions (guide, substrate and superstrate) for a TE or for a TM wave. Physically, if this were <u>not</u> so, eventually a point would be reached along the guide for which a wave within the guide would be unaccompanied by an evanescent wave outside the guide -- clearly an impossibility for a totally internally reflected guided wave.

The optical rays in Figure II-2 represent normals to the propagating, zig-zagging plane wavefronts within the guide. It may be seen that the time it takes for the light to travel along the path $\overline{PQR}$ must be equal to the time it takes for the light (the evanescent wave) in the n_a - region (or, equivalently, in the n_s - region) to travel along the path $\overline{PR}$. Another way of expressing this is to state that the phase delay experienced by the light wave in the guide must be equal (to within an even multiple of π) to the phase delay experienced by the evanescent wave, i.e.,

$$(\text{phase delay})_a + 2m\pi = (\text{phase delay})_g \qquad \text{II-(55)}$$

the evanescent wave's phase delay is:

$$(\text{phase delay})_a = (\overline{PR})\beta = (2W\tan\Theta)\beta \qquad \text{II-(56)}$$

The phase delay in the guide has three contributions, one due to the optical path along $\overline{PQR}$, another due to the phase change

$-2\phi_{ag}$ resulting from total internal reflection from the $n_g - n_a$ boundary, and the third change $-2\phi_{gs}$ due to total internal reflection from the $n_g - n_s$ boundary:

$$(\text{phase delay})_g = (\overline{PQR})\, k_o n_g - 2\phi_{ag} - 2\phi_{gs} \qquad \text{II-(57)}$$

$$= (2W/\cos\theta) k_o n_g - 2\phi_{ag} - 2\phi_{gs} \qquad \text{II-(58)}$$

Substituting Eqs. (56) and (58) into Eq. (55) produces:

$$\beta\, 2W\tan\theta = 2k_o n_g W/\cos\theta - 2\phi_{ag} - 2\phi_{gs} - 2m\pi \qquad \text{II-(59)}$$

Since $\beta = k_o n_g \sin\theta$, Eq. (59) can be written as

$$2\phi_{ag} + 2\phi_{gs} + 2m\pi = 2k_o n_g W/\cos\theta - 2k_o n_g W\sin^2\theta/\cos\theta \qquad \text{II-(60)}$$

$$= (2Wk_o n_g)\left[(1-\sin^2\theta)/\cos\theta\right] \qquad \text{II-(61)}$$

$$= (2Wk_o n_g)\cos\theta \qquad \text{II-(62)}$$

or

$$2Wk_o n_g\cos\theta - 2\phi_{ag} - 2\phi_{gs} = 2m\pi \qquad \text{II-(63)}$$

Eq. (63) is identical with the characteristic waveguide equations (31) and (48), providing that the correct values of ϕ_{ag} and ϕ_{gs} are used.

According to Born and Wolf,[4] the phase changes ϕ_{ag} and ϕ_{gs} for the TE and TM cases of total internal reflection are identical to the values given by Eqs. (29), (30), (44) and (45). Further, once the waveguide equations are known, it is then possible to obtain the waveguide cutoff equations for $(W_{min})_{TE}$ and $(W_{min})_{TM}$. Finally, the extension to TE + TM waves is the same as given above -- by the superposition of TE and TM waves.

Physically, the waveguide equation (63) [i.e., Eqs. (31) or (48)] represents the condition that, after two reflections (one at each boundary), the wavefront is in phase with the original wavefront. If this were not so, many wavefronts, each differing in phase by some amount, might eventually add up somewhere in the guide such that they would destructively interfere. By conservation of energy, the light would then have to pass out into the n_a or n_s regions, i.e., the light could no longer propagate in the n_g - medium.

It may be seen that a quasi-geometrical ray-optics approach can be a useful way of treating wave propagation in an optical waveguide. It provides a means for understanding the physics of the problem at hand; at the same time it results in an accurate representation of the homogeneous-waveguide equation.

CHAPTER III

BASIC PHYSICS OF INHOMOGENEOUS ASYMMETRIC PLANAR OPTICAL WAVEGUIDES

Introduction

An inhomogeneous optical waveguide is one that has a variable index of refraction n_g within the guide. The refractive index profile treated in this dissertation is assumed to lie in the $\hat{x}$-direction only, i.e., transverse to the net direction of propagation. Further, the index profile is assumed to be a monotonic function of x, such that the value of the index is higher near the superstrate. That is, as indicated in Figure III-1, the refractive index of the guide has its maximum at the $n_a - n_g$ boundary and then continuously decreases in the negative $\hat{x}$-direction. The index can be written as

$$n_g = n_g(x) \qquad \text{III-(1)}$$

where

$$n_g(x)\Big|_{x=0} > n_g(x)\Big|_{x=-W} \qquad \text{III-(2)}$$

Quasi-Geometrical Ray-Optics Approach

It is convenient to begin with a quasi-geometrical approach. The justification of the results will be discussed afterwards. Letting β still be the wavevector component along the $\hat{z}$-axis, then β can be written as

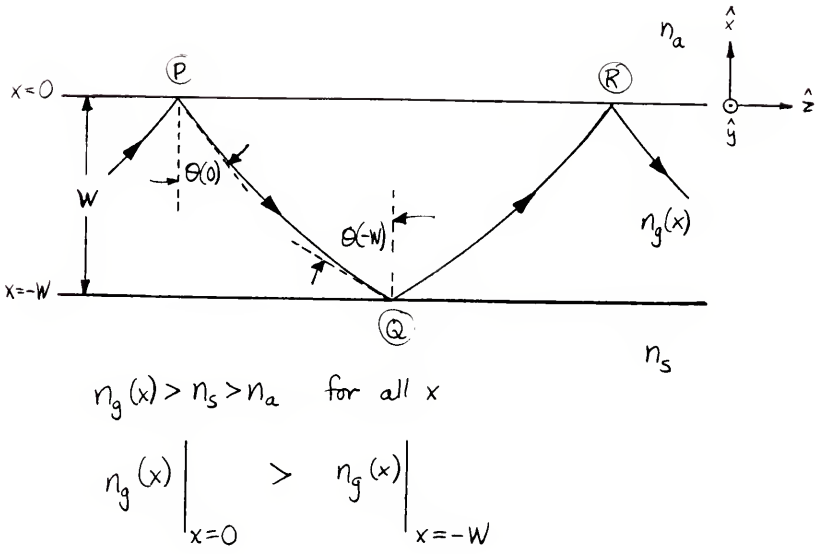


Figure III-1. Cross-sectional view of light propagating in an inhomogeneous asymmetric planar optical waveguide.

$$\beta = k_0 n_g(x) \sin\theta(x) \qquad \text{III-(3)}$$

In Appendix C, solutions were obtained to the wave equation assuming that β was a constant independent of x for both TE and TM waves, i.e., as in the homogeneous case

$$\beta = \text{constant} \qquad \text{III-(4)}$$

$$\equiv k_0 \eta \qquad \text{III-(5)}$$

where

$$\eta \equiv n_g(x) \sin\theta(x) = \text{constant} \qquad \text{III-(6)}$$

Eq. (4) implies that the wave travels at the same velocity v_z in the $\hat{z}$ direction in all three regions: guide, superstrate, and substrate. Eq. (6) is simply another way of writing Snell's Law.

Referring to Figure III-1, it may be seen that as the ray travels from point P to point Q, the ray is constantly bent away from the normal to the $n_a - n_g$ boundary, just as predicted by Eq. (6). Upon reflection at the $n_g - n_s$ boundary, the ray travels from point Q to point R, this time being constantly refracted toward the normal. At the $n_a - n_g$ boundary it is totally internally reflected once again.

In order to obtain the characteristic waveguide equation for the inhomogeneous case, the same argument given in the homogeneous case is useful. That is, the time it takes for the light to travel along the path $\overline{PQR}$ must be equal to the time it takes for the light

(an evanescent wave) in the n_a - region to travel along $\overline{PR}$. As before, this can also be expressed in terms of phase delays:

$$(\text{phase delay})_a + 2m\pi = (\text{phase delay})_g \qquad \text{III-(7)}$$

The evanescent wave's phase delay is

$$(\text{phase delay})_a = (\overline{PR})\beta = (\overline{PR})k_0\eta \qquad \text{III-(8)}$$

The phase delay in the guide has three contributions: one due to the changing optical path caused by the index gradient, $\int \vec{k} \cdot d\vec{r}$, where $r = \hat{x}x + \hat{y}y + \hat{z}z$ ($\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit normals); the other two are due to the reflections at the n_a - n_g and n_g - n_s boundaries: $-2\phi_{ag}$ and $-2\phi_{gs}$. The values of ϕ_{ag} and ϕ_{gs} can be obtained using Eqs. II-(29) and II-(30) for the TE case and Eqs. II-(44) and II-(45) for the TM case. Defining

$$q(x) \equiv n_g(x)\cos\theta(x) \qquad \text{III-(9)}$$

then

$$\phi_{ag}^{TE} = \tan^{-1}\left[a/k_0 q(0)\right] \qquad \text{III-(10)}$$

$$\phi_{gs}^{TE} = \tan^{-1}\left[s/k_0 q(-W)\right] \qquad \text{III-(11)}$$

$$\phi_{ag}^{TM} = \tan^{-1}\left[n_g^2(0)\, a/n_a^2 k_0 q(0)\right] \qquad \text{III-(12)}$$

$$\phi_{gs}^{TM} = \tan^{-1}\left[n_g^2(-W)\, a/n_s^2 k_0 q(0)\right] \qquad \text{III-(13)}$$

where

$$a = k_0 \left(\eta^2 - n_a^2\right)^{1/2} \qquad \text{III-(14)}$$

$$s = k_0 \left(\eta^2 - n_s^2\right)^{1/2} \qquad \text{III-(15)}$$

The phase delay in the guide can thus be written as

$$(\text{phase delay})_g = \int_{\overline{PQR}} \vec{k} \cdot d\vec{r} - 2\varphi_{ag} - 2\varphi_{gs} \qquad \text{III-(16)}$$

Further,

$$\int_{\overline{PQR}} \vec{k} \cdot d\vec{r} = k_0 \int_0^{-W} n_g(x) \cos\theta(x)(-dx) + k_0 \int_{-W}^{0} n_g(x) \cos\theta(x) dx + k_0 \int_{0}^{(\overline{PR})} n_g(x) \sin\theta(x)\, dz \qquad \text{III-(17)}$$

The first integral on the right-hand side of Eq. (17) is due to the $\hat{x}$-component of the optical ray path along $\overline{PQ}$. [The product $\vec{k} \cdot \vec{r}$ is negative, hence the factor (-dx).] The second integral is due to the $\hat{x}$-component of the optical ray path along $\overline{QR}$. [The product $\vec{k} \cdot \vec{r}$ is positive.] The third integral is due to the $\hat{z}$-component of the path along $\overline{PR}$. [The product $\vec{k} \cdot \vec{r}$ is positive also.] Since, by Eq. (6), $n_g(x)\sin\theta(x)$ is a constant, Eq. (17) becomes

$$\int_{\overline{PQR}} \vec{k} \cdot d\vec{r} = 2k_0 \int_{-W}^{0} n_g(x) \cos\theta(x) dx + \left(k_0 \eta\right)(\overline{PR}) \qquad \text{III-(18)}$$

Thus, with the aid of Eqs. (8), (16), and (18), Eq. (7) becomes

$$(k_0\eta)(\overline{PR}) = \left[2k_0 \int_{-w}^{0} n_g(x)\cos\theta(x)\,dx + (k_0\eta)(\overline{PR}) - 2\phi_{ag} - 2\phi_{gs}\right] - 2m\pi \qquad \text{III-(19)}$$

or,

$$2k_0 \int_{-w}^{0} n_g(x)\cos\theta(x)\,dx - 2\phi_{ag} - 2\phi_{gs} = 2m\pi \qquad \text{III-(20)}$$

Eq. (20) is thus the waveguide equation for asymmetric inhomogeneous waveguides, as determined by the quasi-geometrical ray-optics approach. It may be noted that when $n_g(x)$ and $\cos\theta(x)$ are constants [as in the homogeneous case], Eq. (20) reduces to Eqs. II-(31) or II-(48), the waveguide equations for homogeneous waveguides.

Writing

$$\cos\theta(x) = \left[1 - \sin^2\theta(x)\right]^{1/2} \qquad \text{III-(21)}$$

then the integral in Eq. (20) can be transformed to

$$2k_0 \int_{-w}^{0} n_g(x)\left[1-\sin^2\theta(x)\right]^{1/2}dx = 2k_0 \int_{-w}^{0} \left[n_g^2(x) - n_g^2(x)\sin^2\theta(x)\right]^{1/2}dx \qquad \text{III-(22)}$$

$$= 2k_0 \int_{-w}^{0} \left[n_g^2(x) - \eta^2\right]^{1/2}dx \qquad \text{III-(23)}$$

where the last step follows from Eq. (6). Thus Eq. (20) becomes

$$2k_0 \int_{-w}^{0} \left[n_g^2(x) - \eta^2\right]^{1/2}dx - 2\phi_{ag} - 2\phi_{gs} = 2m\pi \qquad \text{III-(24)}$$

which makes it easy to directly incorporate the mathematical form of the index gradient $n_g(x)$. Sometimes the form of $n_g(x)$ allows the integral to be directly integrated analytically. Otherwise (such as with an error function profile) it can be integrated numerically. From Eqs. (10)-(15), it may be seen that Eq. (24) is transcendental in both η and W. Writing

$$\eta = n_g(0) \sin\theta(0) \qquad \text{III-(25)}$$

then, for a given value of $n_g(0)$, perhaps the easiest way to solve Eq. (24) is to pick a value for $\theta(0)$, then try values of W until one is found which satisfies Eq. (24). By using different values of $\theta(0)$, a set of curves can be generated for the different modes m. Also, since $\phi_{ag}^{TM} > \phi_{ag}^{TE}$ and $\phi_{gs}^{TM} > \phi_{gs}^{TE}$, then the value of W obtained from Eq. (24) for the TM case is greater than W obtained for the TE case, as in the homogeneous case.

For the cutoff condition, waves will begin to escape when either

$$\sin\theta(0) = n_a / n_{ag} \qquad \text{III-(26)}$$

or

$$\sin\theta(-W) = n_s / n_{gs} \qquad \text{III-(27)}$$

where

$$n_{ag} \equiv n_g(0) \qquad \text{III-(28)}$$

$$n_{gs} \equiv n_g(-W) \qquad \text{III-(29)}$$

Using Eq. (6), Eqs. (26) and (27) are equivalent to the following equations

$$\eta = n_g(x) \sin\Theta(x) = n_a \qquad \text{III-(30)}$$

$$\eta = n_g(x) \sin\Theta(x) = n_s \qquad \text{III-(31)}$$

respectively. It was assumed for homogenous waveguides that $n_s > n_a$. Hence as Θ decreases from a maximum value of $\pi/2$ (i.e., propagation straight down the guide), then, for a given functional form of n_g, the cutoff condition first to be met is given by Eq. (31). That is, waves will start to leak out at the $n_g - n_s$ boundary when $\sin\Theta(x) = n_s/n_g(x)$ [i.e., when $\sin\Theta(-W) = n_s/n_g(-W)$, or equivalently, when $\sin\Theta(0) = n_s/n_g(0)$]. The waves will still be totally internally reflected at the $n_a - n_g$ boundary. This result is identical with the result obtained for the homogeneous case. Substituting Eq. (31) into Fqs. (10) - (15) produces

$$\phi_{gs}^{TE} = \phi_{gs}^{TM} = 0 \qquad \text{III-(32)}$$

$$\phi_{ag}^{TE} = \tan^{-1}\left[\left(n_s^2 - n_a^2\right)^{1/2} \Big/ \left(n_{ag}^2 - n_a^2\right)^{1/2} \right] \qquad \text{III-(33)}$$

$$\phi_{gs}^{TM} = \tan^{-1}\left[n_g^2(-W) \left(n_s^2 - n_a^2\right)^{1/2} \Big/ n_a^2 \left(n_{ag}^2 - n_a^2\right)^{1/2} \right] \qquad \text{III-(34)}$$

Then at cutoff, Eq. (24) becomes for the TE case

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_s^2\right]^{1/2} dx = \frac{\lambda_0}{2\pi} \left\{ m\pi + \tan^{-1}\left[\frac{\left(n_s^2 - n_a^2\right)^{1/2}}{\left(n_{ag}^2 - n_s^2\right)^{1/2}} \right] \right\} \qquad \text{III-(35)}$$

and, for the TM case,

$$\int_{-(W_{min})_{TM}}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \frac{\lambda_0}{2\pi}\left\{m\pi + \tan^{-1}\left[\frac{n_{ag}^2}{n_a^2}\frac{(n_s^2-n_a^2)^{1/2}}{(n_{ag}^2-n_s^2)^{1/2}}\right]\right\} \quad \text{III-(36)}$$

As mentioned above, Eqs. (35) and (36) hold when $n_s > n_a$. However, for the inhomogeneous case, the situation $n_s < n_a$ is different than $n_s > n_a$ [since $n_g(0) \neq n_g(-W)$]. Hence if $n_s < n_a$, then the cutoff condition is given by Eq. (30), in which case $\phi_{gs}^{TE} \neq 0$, $\phi_{gs}^{TM} \neq 0$, $\phi_{ag}^{TE} = \phi_{gs}^{TM} = 0$, and Eq. (24) becomes

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \frac{\lambda_0}{2\pi}\left\{m\pi + \tan^{-1}\left[\frac{(n_a^2-n_s^2)^{1/2}}{(n_{gs}^2-n_a^2)^{1/2}}\right]\right\} \quad \text{III-(37)}$$

and

$$\int_{-(W_{min})_{TM}} \left[n_g^2(x) - \eta^2\right]^{1/2} dx = \frac{\lambda_0}{2\pi}\left\{m\pi + \tan^{-1}\left[\frac{n_{gs}^2}{n_s^2}\frac{(n_a^2-n_s^2)^{1/2}}{(n_{gs}^2-n_a^2)^{1/2}}\right]\right\} \quad \text{III-(38)}$$

It may be seen that, indeed, the substitution of n_a for n_s and vice versa in Eqs. (37) and (38) does not result in the same relations as given by Eqs. (35) and (36) because $n_{ag} \neq n_{gs}$, unlike the homogeneous case.

Since $\sin\Theta(0)$ cannot exceed a value of unity, then from Eq. (6) [or Eq. (25)], η can never exceed the value $n_g(0)$. In fact, η cannot equal $n_g(0)$ either, as may be seen from the following argument. If $\sin\Theta(0) = 1$, then $\Theta(0) = \pi/2$, and from Snell's Law [Eq. (6)] $\Theta(x) = \pi/2$ for all x. Hence the integrand in Eq. (20) would be zero for all x, and in order to propagate a light wave, only an infinitely thick ($W \to \infty$) guide could satisfy Eq. (20), which is physically unrealistic. Thus, using Eq. (25), the range of values of $\Theta(0)$ for

waveguiding in the case $n_s > n_a$ is given by the relationship

$$n_g > \eta > n_s \qquad \text{III-(39)}$$

and for the $n_a > n_s$ case,

$$n_g > \eta > n_a \qquad \text{III-(40)}$$

Eqs. (39) and (40) apply to both TE and TM modes as determined by Eq. (24). It may be seen that Eq. (39) is identical to Eq. II-(35) for the homogeneous case.

Solutions for $\vec{E}$ and $\vec{H}$

As in Appendix B, in Appendix C trial solutions are chosen for the vector potentials $\vec{F}$ and $\vec{A}$, only now a provision is made in the form of the solution to allow for an index gradient within the guide. Specifically, instead of the product gx in Eqs. II-(24) and II-(42), where g is a constant given by Eq. II-(27), the choice is g = g(x). The function $g(x)/k_o$ is often called the eikonal.[5] Using the boundary conditions[6] on $\vec{E}$ and $\vec{H}$ together with the assumption that the index in the guide is a slowly varying function of x, the following solutions are obtained in Appendix C for TE modes:

$$E_{oy}(x) = \begin{cases} 2\,g^{-1/2}(0)\,F_2\,e^{-i\phi_{ag}^{TE}}\cos\left(\phi_{ag}^{TE}\right)e^{-ax} & \text{for } x \geqq 0 \qquad \text{III-(41)} \\ 2\,g^{-1/2}(x)\,F_2\,e^{-i\phi_{ag}^{TE}}\cos\left[g(x)+\phi_{ag}^{TE}\right] & \text{for } -W \leqq x \leqq 0 \qquad \text{III-(42)} \\ 2\,g^{-1/2}(-W)\,F_2\,e^{-i\phi_{ag}^{TE}}\cos\left(\phi_{gs}^{TE}+m\pi\right)e^{s(W+x)} & \text{for } x \leqq -W \qquad \text{III-(43)} \end{cases}$$

where Eqs. II-(17) through II-(21) apply equally well to the TE inhomogeneous case. The exact form of H_{ox} and H_{oz} is given in Eqs. C-(135) through C-(138). It may be seen from Eq. (42) that both the amplitude and phase of $E_{oy}(x)$ are functions of x. The function g(x) is given by

$$g(x) = k_o \int_0^x q(x)\, dx \qquad \text{III-(44)}$$

where q(x) is given by Eq. III-(9). The method used in Appendix C for solving the wave equation for an inhomogeneous medium is called the WKBJ method[7-10] (after Wentzel, Kramers, Brillouin, and Jeffries). The assumption that the guide has a slowly varying index profile leads to the follwing requirement:

$$\frac{1}{k_o^2} \left| 3\left(\frac{q'}{2q^2}\right)^2 - \frac{q''}{2q^3} \right| << 1 \qquad \text{III-(45)}$$

where q' and q'' are the first and second derivatives of q with respect to x. Eq. (45) is the WKBJ approximation for the TE case.[10] It may be seen that for large wavenumbers k_o (i.e., small λ_o), the WKBJ approximation is more easily satisfied. However, as $q = n_g \cos\Theta$ approaches zero (as n_g decreases and Θ approaches $\pi/2$), Eq. (45) cannot be satisfied, regardless of how small q' and q'' are. This situation will be discussed in Chapter IV. When Eq. (45) <u>is</u> satisfied, all of the TE results obtained from the quasi-geometric ray-optics approach are valid, at least to the level of approximation given by Eq. (45). In the homogeneous case q is a constant. Hence, the function g(x) in Eq. (44) can be easily evaluated to give $(k_o q)x = (k_o n_g \cos\Theta)x$, which

is the same as the product gx originally obtained in Chapter II, as mentioned above.

For the TM case, the following solutions are obtained in Appendix C:

$$H_{oy}(x) = \begin{cases} -2n_g(0)\, q^{-1/2}(0)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{ag}^{TM}\right) e^{-ax} & \text{for } x \geqq 0 \quad \text{III-(46)} \\ -2n_g(x)\, q^{-1/2}(x)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left[g(x) + \varphi_{ag}^{TM}\right] & \text{for } -W \leqq x \leqq 0 \quad \text{III-(47)} \\ -2n_g(-W)\, q^{-1/2}(-W)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{gs}^{TM} + m\pi\right) & \text{for } x \leqq -W \quad \text{III-(48)} \end{cases}$$

where Fqs. II-(36) through II-(40) apply to the TM inhomogeneous case. The exact form of E_{ox} and E_{oz} is given in Eqs. C-(224) through C-(229). It may be noted that the amplitude dependence of H_{oy} on x has two contributions, $n_g(x)$ and $q^{-1/2}(x)$, whereas for the TE amplitude dependence only one factor, $q^{-1/2}(x)$, is present. The WKBJ approximation for the TM case[10] is given by

$$\frac{1}{k_o^2} \left| \left\{ 3\left(\frac{q'}{2q^2}\right)^2 - \frac{q''}{2q^3} \right\} + \frac{1}{q^2} \left\{ \frac{n_g''}{n_g} - 2\left(\frac{n_g'}{n_g}\right)^2 \right\} \right| \ll 1 \qquad \text{III-(49)}$$

which has the additional terms (given in the second brace) due to the first and second derivatives of n_g. The situation when Eq. (49) is not satisfied will be discussed in Chapter IV. When Eq. (49) is satisfied, all of the TM results obtained from the quasi-geometric method are valid, at least to the level of approximation given by Eq. (49).

CHAPTER IV

BASIC PHYSICS OF TOTAL INTERNAL REFRACTION IN INHOMOGENEOUS PLANAR OPTICAL WAVEGUIDES

Introduction

It was stated in Chapter III (and shown in Appendix C) that the WKBJ solutions for $\vec{E}$ and $\vec{H}$ are valid as long as Eqs. III-(45) and III-(49) for TE and TM modes respectively are satisfied. When either Eq. III-(45) or III-(49) is not satisfied, it may be seen from Eqs. III-(42), C-(135), III-(47), and C-(225) that, due to the factor $q^{-1/2}$, at least two field components (E_{oy} and H_{ox} for the TE case, H_{oy} and E_{ox} for the TM case) become infinite as $q(x) = n_g(x) \cos\Theta(x)$ approaches zero as $\Theta(x) \to \pi/2$. Since this is physically unrealistic, some other approach must be used to solve the wave equation near the level $q = 0$. In the field of radio ionospheric physics, Budden[11] and Wait[12] have developed a useful approach for describing the propagation of light near $q = 0$. This method will be adapted to optical waveguides. Before doing this, however, it is convenient to use quasi-geometrical arguments to form a physically based explanation of the results obtained in the more rigorous derivation. This permits a practical understanding of the phenomenon of total internal refraction-and-reflection in general-index-gradient planar optical waveguides.

Quasi-Geometrical Ray-Optics Approach

First of all, it should be mentioned that -W is now chosen to be

the value of x for which $\Theta = \pi/2$ (and $q = 0$) for a given mode m and incident angle $\Theta(0)$, as indicated in Figure IV-1. That is, at the level $x = -W$, $\Theta(-W) = \pi/2$, $q(-W) = 0$, and total internal refraction occurs. It is convenient to let

$$n_g(x) = n_o + b(x)\Delta n \qquad \text{IV-(1)}$$

where

$$n_g(0) = n_o + \Delta n \qquad \text{IV-(2)}$$

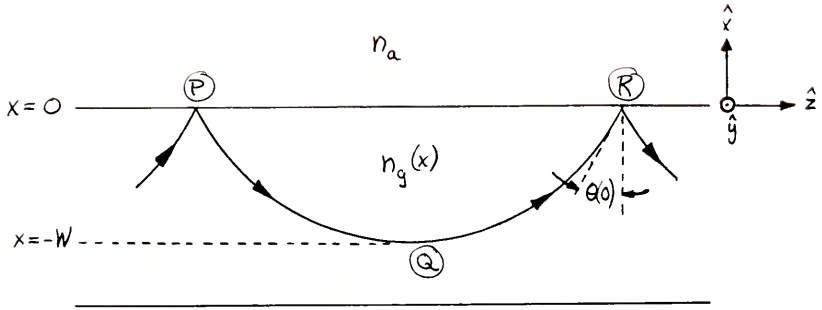
at the top of the guide, and

$$n_g(-D) = n_o \qquad \text{IV-(3)}$$

where -D is the penetration depth of the index profile such that beyond $x = -D$ there is no longer an index gradient. For instance, n_o might be the index of the film before ion implanation, and Δn the maximum index change (where $\Delta n << n_o$) which is generally a measurable quantity. The function b(x) thus gives the index profile, where

$$b(0) = 1 \qquad \text{IV-(4)}$$

$$b(-D) = 0 \qquad \text{IV-(5)}$$



$n_g(x) > n_a$ for all x

$n_g(0) > n_g(-W)$

$\theta(-W) = \pi/2$

Figure IV-1. Cross-sectional view of light propagating in an inhomogeneous planar optical waveguide by means of total internal refraction and reflection.

As shown in Figure IV-1, after reflecting from the $n_a - n_g(0)$ boundary, the wave is continually refracted away from the normal to the boundary until, at the level $x = -W$ (and $\Theta = \pi/2$), the wave is refracted back toward the normal as it returns toward the $n_a - n_g(0)$ boundary. The phase change suffered by this wave has three contributions: a) the integral of the changing optical path $\vec{k} \cdot \vec{dr}$ over the path $\overline{POR}$; b) the phase change $-2\phi_{ag}$ upon total internal reflection at the $n_a - n_g(0)$ boundary; and c) the phase change upon total internal refraction. It is the third contribution which is the more elusive quantity.

From one point of view, since the total internal refraction phenomenon is purely refraction, no phase change might be expected. On the other hand, since the light does in fact reverse its direction, then a phase change of π might seem reasonable, since in the limit as $\Theta \to \pi/2$, the phase change $2\phi_{ag}$ due to total internal reflection approaches π. However, neither of these points of view is correct. In fact, the phase change is $\pi/2$, not 0 or π.

It is known[13] that a phase change of π occurs when light passes through the focus of a spherical lens; for a cylindrical lens, there is a phase change of $\pi/2$. In Figure IV-2a, it may be seen that as a beam of light is refracted at a boundary with an abrupt refractive index change, the beam width decreases when $n_2 < n_1$. For a continuously varying index, a similar phenomenon occurs, as shown in Figure IV-3. Here the path of the ray is shown surrounded by a small beam width. Because the waveguide is a planar slab (as opposed to a fiber), it may be seen that in the region $x = -W$ and $(q = 0)$, the beam undergoes a cylindrical focus. Hence, it is reasonable to expect a phase change of $\pi/2$ upon "reflection", i.e., total internal refraction, at this level.

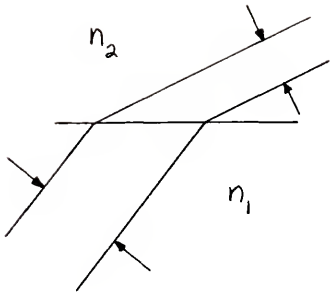


Figure IV-2a

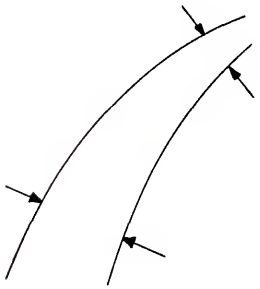


Figure IV-2b

Figure IV-2. Refraction of light at a boundary with an abrupt change of refractive index (Figure IV-2a) and in a medium with a continuously varying index (Figure IV-2b).

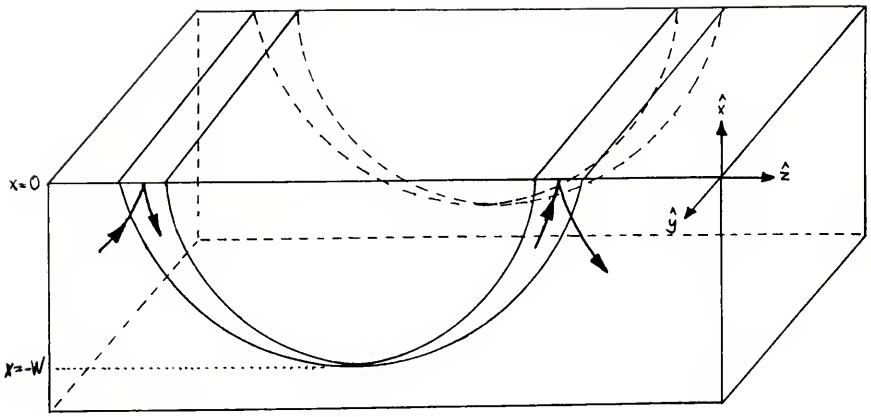


Figure IV-3. Total internal refraction in a planar optical waveguide showing the cylindrical focussing effect.

From these various phase contributions the characteristic waveguide equation is thus obtained:

$$2k_o \int_{-W}^{0} n_g(x) \cos\theta(x)\, dx - 2\phi_{ag} - \frac{\pi}{2} = 2m\pi \qquad \text{IV-(6)}$$

which applies to both TE and TM modes. [A second-order correction for TM modes will be discussed later.] Comparison of Eq. (6) with Eq. III-(20) shows that the only change is to replace the term $2\phi_{gs}$ with $\pi/2$. Eq. (6) can also be written as

$$2k_o \int_{-W}^{0} [n_g^2(x) - \eta^2]^{1/2} dx - 2\phi_{ag} - \frac{\pi}{2} = 2m\pi \qquad \text{IV-(7)}$$

similar to Eq. III-(24). Eqs. (6) and (7) are transcendental in η, which is equivalent to saying that it is transcendental in $\theta(0)$, according to Eq. III-(25). It may or may not be transcendental in W too, depending on the functional form of $n_g(x)$ and the integral in Eq. (7). Eqs. (6) and (7) are useful for describing the propagation of light in such optical waveguides and has general applicability to any slowly varying monotonic index profile (including the error function profile); it also allows computation of numerical answers. As before, since $\phi_{ag}^{TM} > \phi_{ag}^{TE}$, the value of W for the TM case will exceed W for the TE case. It is interesting to note that, in the limit as $\theta(0) \to \pi/2$, Eq. (7) approaches an equation for TE modes derived by Marcuse.[14]

Following arguments given in Chapter III, the cutoff condition occurs when either [from Eq. III-(30)]

$$\eta = n_a \qquad \text{IV-(8)}$$

or [similar to Eq. III-(31)]

$$\eta = n_o \qquad \text{IV-(9)}$$

For $n_o > n_a$ (which is usually the case), the cutoff condition is given by Eq. (9), i.e., when $\eta = n_o$. When $\eta < n_o$, the light can no longer be totally internally refracted, and the light escapes below the guide. For this cutoff condition, there is thus no $\pi/2$ phase change. For the exact form of the phase changes $2\phi_{ag}^{TE}$ and $2\phi_{ag}^{TM}$, Eq. (9) can be substituted into Eqs. III-(10), III-(12), and III-(14). Further, from Eqs. III-(6) and III-(9), $q(x)$ can be written as

$$q(x) = \left[n_g^2(x) - \eta^2\right]^{1/2} \qquad \text{IV-(10)}$$

Using Eqs. (2), (9), and (10), then

$$q(0) = (2 n_o \Delta n)^{1/2} \qquad \text{IV-(11)}$$

where the $(\Delta n)^2$ term has been neglected compared to $2n_o\Delta n$. Thus, at cutoff the phases are

$$\phi_{ag}^{TE} = \tan^{-1}\left[\frac{(n_o^2 - n_a^2)^{1/2}}{(2 n_o \Delta n)^{1/2}}\right] \qquad \text{IV-(12)}$$

$$\phi_{ag}^{TM} = \tan^{-1}\left[\frac{n_o^2 + 2n_o\Delta n}{n_a^2}\ \frac{(n_o^2 - n_a^2)^{1/2}}{(2 n_o \Delta n)^{1/2}}\right] \qquad \text{IV-(13)}$$

Eq. IV-(7) then becomes an integral equation for W_{min}:

$$\int_{-(W_{min})}^{0} \left[n_g^2(x) - n_o^2\right]^{1/2} dx = \frac{\lambda_o}{2\pi}(m\pi + \phi_{ag}) \qquad \text{IV-(14)}$$

where for $(W_{min})_{TE}$ Eq. (12) is to be used in Eq. (14), and for $(W_{min})_{TM}$ Eq. (13) is to be used.

In the event that $n_o < n_a$, a much less common situation [where $n_a < n_o + \Delta n$, in order to insure total internal reflection at the $n_a - n_g(0)$ boundary], then the cutoff condition is given when $\eta = n_a$ [Eq. (8)], and the light leaks out into the superstrate. In this case $\phi_{ag} = 0$. Thus Eq. (7) becomes

$$\int_{-(W_{min})}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \lambda_o (4m+1)/8 \qquad \text{IV-(15)}$$

which is the same for both TE and TM modes (except for the second-order correction for the TM case, to be discussed).

As explained in Chapter III, η can never exceed or equal the value $n_g(0)$. Thus the range of values of $\Theta(0)$ for waveguiding without cutoff in the case $n_s > n_a$ is given by the relationship

$$n_o + \Delta n > \eta > n_o \qquad \text{IV-(16)}$$

and for $n_o < n_a$

$$n_o + \Delta n > \eta > n_a \qquad \text{IV-(17)}$$

Since $\Theta(-W) = \pi/2$, then from Eqs. (9) and III-(6), $\eta = n_g(-W) = n_o$. Referring to Eq. (1), this can be satisfied only when $b(-W) = 0$, which from Eq. (5) shows that, at cutoff, W equals the maximum index gradient penetration depth D.

It was mentioned above that there is a second-order correction to the characteristic waveguide equation for TM modes. Physically, the reason for this is as follows. In the TE case, the only component of $\vec{E}_{TE}$ is E_y, which is transverse to the index gradient. Hence, as a TE wave propagates within the guide, the $\vec{E}$-vector at any one point, say x_o, within the guide will "see" only one index -- $n_g(x_o)$ -- at a

time. For the TM case, however, since there is a component of $\vec{E}$ along the index gradient (the E_x component), the situation becomes more complicated. Thus when deriving the $\pi/2$ phase change for the characteristic waveguide equation for the TM case, it was implicitly assumed that the TM wave "sees" some sort of average index $n_g(x_o)$ at at any given position x_o on the ray path. As shown below, this averaging takes place over some fraction of a wavelength of light, say $C_1\lambda(x_o)$, where C_1 is a constant to be determined and where $\lambda(x) = \lambda_o/n_g(x)$. Any shift δW in the level of total internal refraction at $x = -W$ can then be considered as due to the change of wavelength $\Delta\lambda$ caused by a different refractive index above and below the ray path. That is, the index above the ray path, say $n_g^+(-W)$, is slightly greater than $n_g(-W)$, and the index below the path, say $n_g^-(-W)$, is slightly less than $n_g(-W)$. Thus the wavelength λ^+ slightly above the ray path is smaller than $\lambda(-W)$, and the wavelength λ^- slightly below the ray path is slightly larger than $\lambda(-W)$. Hence the median position of λ (over which E_x is averaged) occurs at a level slightly <u>below</u> $x = -W$. The net effect is that the E_x component sees the index $n_g(-W)$ <u>before</u> the ray actually reaches the level $x = -W$. As a result, the TM ray is internally refracted "sooner" than in the TE case, i.e., at a value of $|x|$ which is slightly <u>smaller</u> than in the TE case. This has a slight tendency to offset the fact that the value of W for the TM case (from the waveguide equation) is larger than W for the TE case.

It is possible to derive the functional form of this shift δW. From the above argument, it is reasonable to expect that δW is proportional to the change in wavelength $\Delta\lambda$ due to the slightly different refractive indexes n_g^+ and n_g^-, i.e.,

$$\delta W = C_2 \Delta\lambda \qquad \text{IV-(18)}$$

where C_2 is another constant. Since $\lambda = \lambda_o/n_g$, then

$$\Delta\lambda = \frac{-\lambda_o}{n_g^2(-W)} \Delta n_g \qquad \text{IV-(19)}$$

Further, for an index gradient dn_g/dx near the level $x = -W$, the change in index seen over a distance $C_1\lambda$ is

$$\Delta n_g = C_1 \lambda \left.\frac{dn_g}{dx}\right|_{x=-W} \qquad \text{IV-(20)}$$

$$= C_1 \lambda_o \left.\left(\frac{1}{n_g(x)} \frac{dn_g}{dx}\right)\right|_{x=-W} \qquad \text{IV-(21)}$$

Combining Eqs. (18) - (21) produces

$$\delta W = -C_1 C_2 \lambda_o^2 \left.\left(\frac{1}{n_g^3(x)} \frac{dn_g}{dx}\right)\right|_{x=-W} \qquad \text{IV-(22)}$$

Using Eq. (1)

$$\left.\frac{dn_g}{dx}\right|_{x=-W} = \Delta n\, b'(-W) \qquad \text{IV-(23)}$$

$$n_g^3(-W) = n_o^3 + 3n_o^2 \Delta n + 3n_o(\Delta n)^2 + (\Delta n)^3 \qquad \text{IV-(24)}$$

Since $\Delta n << n_o$, then Eq. (22) becomes approximately

$$\delta W \cong -C_1 C_2 \lambda_o^2 \frac{\Delta n\, b'(-W)}{n_o^3} \qquad \text{IV-(25)}$$

Writing $\lambda_o^2 = 4\pi^2/k_o^2$, then

$$\delta W \cong -C_1 C_2 \, 4\pi^2 \, \frac{\Delta n \, b'(-W)}{n_0^3 k_0^2} \qquad \text{IV-(26)}$$

The product C_1C_2 is evaluated in the more rigorous derivation given in Appendix D and has the value $3/8\pi^2$ [from Eq. D-(122)]. Thus

$$\delta W = -\frac{3}{2} \, \frac{\Delta n \, b'(-W)}{n_0^3 k_0^2} \qquad \text{IV-(27)}$$

In order to get a rough estimate of how large δW might be, if $\Delta n = 0.1$, $D = 1\mu m$, and $W = D$ (i.e., the wave is propagating nearly at cutoff), then, for a linear index profile, $b'(-W) = 0.1\mu m^{-1}$, which is a fairly steep profile. Further, letting $\lambda_o = 0.63\mu m$ and $n = 1.5$, then $\delta W \approx 4.5 \times 10^{-5}\mu m = 0.45$ Å, which is very small. For less steep profiles, δW will be smaller still. Thus to a very good approximation δW can be neglected in calculations concerning the TM characteristic waveguide equation as well as the cutoff thickness $(W_{min})_{TM}$.

Solutions for $\vec{E}$ and $\vec{H}$

1. The Linear Approximation

As mentioned in the introduction of this chapter, the WKBJ solutions for the field components E_{oy} and H_{ox} (for the TE case) and H_{oy} and E_{ox} (for the TM case) become infinite as $q(x)$ approaches zero. In order to get around this problem, theoretical techniques developed by Budden[11] and Wait[12] for radio ionospheric physics can be adapted to optical waveguides.

The assumption is now made that, for a slowly varying index profile, q^2 is approximately linearly proportional to x near the level $q = 0$, i.e., near $x = -W$. This is nearly the same thing as saying that $n_g(x)$ is linear in this region, as may be seen from the following argument.

Expanding $n_g(x)$ in a Taylor series expansion around $x = -W$, produces

$$n_g(x) = n_g(-W) + (x+W)\, n_g'(-W) + \cdots \qquad \text{IV-(28)}$$

which, from Eq. (1), becomes

$$n_g(x) = n_o + \Delta n\left[b(-W) + b'(-W)(x+W)\right] \qquad \text{IV-(29)}$$

which is linear in x. Further, for small Δn

$$n_g^2(x) \cong n_o^2 + 2n_o\,\Delta n\left[b(-W) + b'(-W)(x+W)\right] \qquad \text{IV-(30)}$$

which is also linear in x. At the level $q = 0$ (and $x = -W$), $\Theta(-W)$ has the value $\pi/2$, so from Eq. III-(6)

$$\eta^2 = n_g^2(-W)\sin^2\Theta(-W) \qquad \text{IV-(31)}$$

$$= n_g^2(-W) \qquad \text{IV-(32)}$$

Thus, from Eq. (30),

$$\eta^2 = n_o^2 + 2n_o\,\Delta n\, b(-W) \qquad \text{IV-(33)}$$

hence [from Eq. (10)]

$$q^2(x) = n_g^2(x) - \eta^2 = 2n_o\,\Delta n\, b'(-W)(x+W) \qquad \text{IV-(34)}$$

which is linear in x too. Eq. (34) is equivalent to a Taylor expansion of $q^2(x)$ in the region near $x = -W$.

2. TE Modes

Defining the following quantity:[11]

$$\zeta \equiv -\left[\frac{k_o}{(q^2)'}\right]^{2/3} q^2 \qquad \text{IV-(35)}$$

where $(q^2)'$ is the derivative of q^2 [from Eq. (34)] with respect to x, then, as shown in Appendix D, the wave equation for F_x [Eq. A-(92), or equivalently, Eq. D-(20)] can be written as

$$\frac{d^2 F_x}{d\zeta^2} - \zeta F_x = 0 \qquad \text{IV-(36)}$$

which is called the Stokes differential equation. Its solution as discussed by Budden[11] is the following:

$$F_x = K_{TE}\, e^{j(\omega t - \beta z)}\, Ai(\zeta) \qquad \text{IV-(37)}$$

where K_{TE} is a constant and where $Ai(\zeta)$ is the Airy integral function defined as follows:

$$Ai(\zeta) = \frac{1}{\pi}\int_0^\infty \cos\left(\zeta u + \tfrac{1}{3}u^3\right) du \qquad \text{IV-(38)}$$

Tables of $Ai(\zeta)$ are given by Miller.[15] A plot of $Ai(\zeta)$ versus ζ is given in Figure IV-4. It may be seen that beyond the region $\zeta = 0$ ($q = 0$, $x = -W$) the solution $Ai(\zeta)$ is an exponentially damped wave -- an evanescent wave.

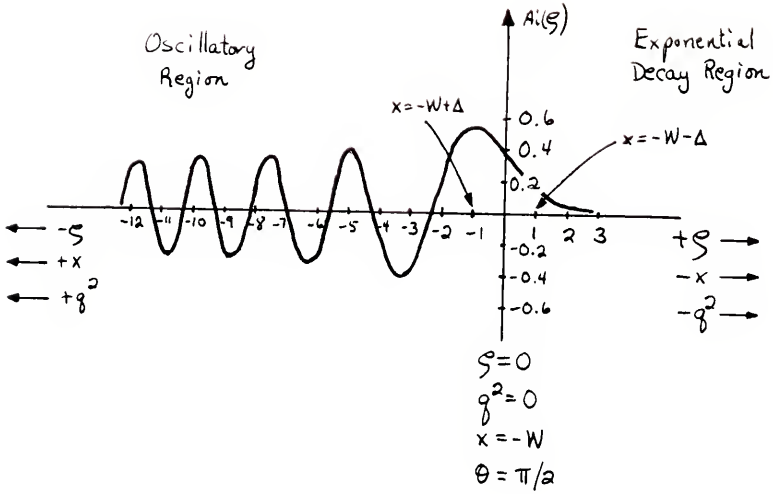


Figure IV-4. Plot of the Airy integral function Ai(ζ) as a function or ζ, or, equivalently, as a function of q^2.[11,12]

There is another solution to Eq. (26) since it is a second-order differential equation. It is another Airy integral function $Bi(\zeta)$. However, beyond the level $x = -W$ at $q = 0$, the function $Bi(\zeta)$ becomes indefinitely large[11] as ζ increases, i.e., as $x \to -\infty$. Since it is physically unrealistic for F_x (and thus $\vec{E}$) to get larger and larger at distances farther and farther away from the waveguide, then the solution to Eq. (26) cannot contain a multiple of $Bi(\zeta)$.

One of the interesting properties of $Ai(\zeta)$, or equivalently $Ai(q^2)$, is that for large q (corresponding to large $\cos\Theta$, far away from the level $q = 0$), $Ai(\zeta)$ asymptotically approaches the WKBJ solution for $\vec{E}$, as shown by Budden.[11] Referring to Figure IV-5 and following the arguments of Budden[11] and Wait,[12] it is assumed that q^2 is linear over the entire region be. There is a certain range cd where $|q|$ is so small that it violates the WKBJ approximation, so the WKBJ solutions cannot be used. Outside this range the WKBJ solutions are good approximations. In the regions bc and de the asymptotic approximations of $Ai(\zeta)$ can be used, and they must be matched to the WKBJ solutions. In other words, outside the region near $q = 0$, the WKBJ solutions should be valid; inside the region, $Ai(\zeta)$ must be used. At the "boundaries" of this region, at say $x = -W \pm \Delta$ (where Δ is to be determined), the two solutions must merge smoothly, thus providing boundary conditions to be satisfied. When this is done, as shown in Appendix D, the phase change δ_R suffered by the wave upon propagating into the linear n_g (i.e., linear q^2) region, reflecting (i.e., internally refracting) at the level $x = -W$, and then propagating out of the linear n_g region, is given by

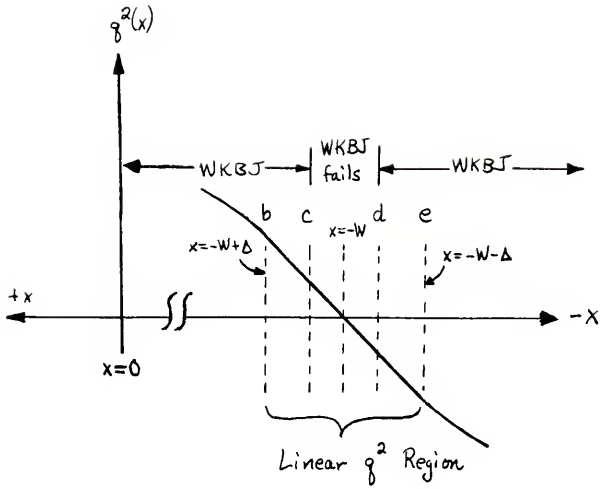


Figure IV-5. Plot of $q^2(x)$ as a function of x showing the linear region be.[11,12]

$$e^{i\delta_R} = i\, e^{-2ik_0 \int_{-W+\Delta}^{-W} q\, dx} \qquad \text{IV-(39)}$$

Aside from the factor $i = e^{i\pi/2}$, this phase change δ_R is nothing more than what would be predicted by the WKBJ method even though the WKBJ solution itself breaks down in the region $q = 0$. When δ_R is combined with the phase change suffered by the wave in the rest of the guide, the characteristic waveguide equation [Eq. (6)] results.

The effect of curvature of the index profile must be investigated in order to determine the validity of Eq. (34) and the resultant conversion of the wave equation to the Stokes equation. If an additional term is kept in Eq. (29), then

$$n_g^2(x) \cong n_0^2 + 2n_0\Delta n\left[b(-W) + b'(-W)(x+W) + b''(-W)(x+W)^2/2\right] \qquad \text{IV-(40)}$$

and

$$\eta^2 = n_g^2(-W) = n_0^2 + 2n_0\Delta n\, b(-W) \qquad \text{IV-(41)}$$

Then, from Eq. (10),

$$q^2 = (x+W)B_1 + (x+W)^2 B_2 \qquad \text{IV-(42)}$$

where

$$B_1 \equiv 2n_0\Delta n\, b'(-W) \qquad \text{IV-(43)}$$

$$B_2 \equiv n_0\Delta n\, b''(-W) \qquad \text{IV-(44)}$$

When Eq. (42) is substituted into the wave equation for F_x [Eq. A-(92)], then as shown in Appendix D the wave equation can be converted to the Stokes equation for which Ai(ζ) is solution only if

$$|\zeta| >> |\zeta^2 (k_0 B_1^2)^{-2/3} B_2| \qquad \text{IV-(45)}$$

and this must hold for values of ζ large enough for the asymptotic approximation of Ai(ζ) (and thus the WKBJ solutions) to be valid. According to Budden,[11] this will happen when $|\zeta| \gtrsim 1$. For $|\zeta| = 1$, Eq. (44) requires the following linearity criterion:

$$|B_2| << |k_0 B_1^2|^{2/3} \qquad \text{IV-(46)}$$

which from Eqs. (1), (43), and (44), requires that

$$n_g''(-W) << (16 n_0 k_0^2)^{1/3} [n_g'(-W)]^{4/3} \qquad \text{IV-(47)}$$

The condition that $|\zeta| = 1$ (at $x = -W \pm \Delta$, the boundaries of the linear n_g region, as shown in Figures IV-4 and IV-5), corresponds to

$$\Delta = |x + W| = \frac{1}{[2k_0^2 n_0 n_g'(-W)]^{1/3}} \qquad \text{IV-(48)}$$

It is now possible to write down the solutions for $\vec{E}_{TE}$ and $\vec{H}_{TE}$. The solutions for $x \geq 0$ are given by Eqs. III-(41), C-(135), and C-(136). For the region $(-W + \Delta) \leq x \leq 0$, the solutions are the WKBJ solutions given by Eqs. III-(42), C-(135), and C-(137). For the region $(-W - \Delta) \leq x \leq (-W + \Delta)$, the solutions involve the Airy integral function:

$$E_{oy}(x) = i\beta K_{TE}\ Ai(S) \qquad \text{IV-(49)}$$

where K_{TE} is defined in Eq. D-(58) [using Eq. D-(48)]. The exact form of H_{ox} and H_{oz} is given in Eqs. D-(94) and D-(96). Finally, for the region $x \leq (-W - \Delta)$, the WKBJ exponentially decaying solution is

$$E_{oy}(x) = -\,|g|^{-1/2} F_4\, e^{-k_o \int_W^x |g|\, dx} \qquad \text{IV-(50)}$$

where F_4 is defined in Eq. D-(62). The exact form of H_{ox} and H_{oz} in this region is given in Eqs. D-(99) and D-(100).

3. TM Modes

Defining the following quantities:[7,11]

$$U(x) \equiv A_x / n_g \qquad \text{IV-(51)}$$

$$Q^2(x) \equiv q^2 - \frac{3}{4k_o^2 n_g^4}\left[\frac{\partial(n_g^2)}{\partial x}\right] + \frac{1}{2k_o^2 n_g^2}\frac{\partial^2(n_g^2)}{\partial x^2} \qquad \text{IV-(52)}$$

(where Q is an "effective value of q"),[11] then, as shown in Appendix D, the wave equation for A_x [Eq. A-(93)] can be simplified to

$$\frac{\partial^2 U}{\partial x^2} + k_o^2 Q^2(x)\, U = 0 \qquad \text{IV-(53)}$$

similar to Eq. D-20. By making the assumption that $n_g(x)$ varies linearly near the level $x = -W$, then it is shown in Appendix D that the difference between Q^2 and q^2 from Eq. (52) is proportional to $[\Delta n\, b'(-W)/k_o]^2$, which is a very small quantity for optical waveguides. Hence, near the level $x = -W$, the wave equation Eq. (53) becomes approximately

$$\frac{\partial^2 U}{\partial x^2} + k_o^2 q^2(x) U = 0 \qquad \text{IV-(54)}$$

and all the techniques described for the TE case can be used to convert this form of the wave equation to the Stokes equation, and to then find the Airy integral solutions.

Neglecting the difference between Q^2 and q^2 is equivalent to the assumption that they have a zero at nearly the same point. As discussed in the quasi-geometrical approach, it is possible to calculate this (very) slight difference δW between the TE and TM reflection points (i.e., the points of total internal refraction). Since δW is so small, Eq. (54) is valid for TM modes to a very good approximation.

The quantity $Q^2(x)$ is subject to two linearity criterions, as compared to just one [Eq. (46)] for $q^2(x)$. Assuming Q^2 obeys the criterion given by Eq. (46) concerning the second derivative $n_g''(-W)$, then a Taylor series expansion of Q^2 from Eq. (52) can still result in a quadratic term which has the first derivative $n_g'(-W)$ only [Eq. D-(123)]. As a result [from Eq. D-(128)], Q^2 may be written as

$$Q^2(x) = (x+W) Q_1 - (x+W)^2 Q_2 \qquad \text{IV-(55)}$$

where δW has been neglected and where the constants Q_1 and Q_2 are defined in Eqs. D-(120) and D-(126) respectively. Eq. (55) is of the same form as Eq. (42); thus the quadratic term in Eq. (55) is also subject to the linearity criterion given by Eq. (46), i.e.,

$$|Q_2| << |k_o Q_1^2|^{2/3} \qquad \text{IV-(56)}$$

Using the definitions of Q_1 and Q_2 together with the fact that the minimum value of η is n_o before cutoff, Eq. (56) requires that

$$\eta^2 >> 6.4\ \Delta n\ b'(-W)/k_o \qquad \text{IV-(57)}$$

Budden[11] defines the quantity B as follows:

$$B \equiv \eta^2 \left\{ k_o / \left[n_g^2(x) \right]' \Big|_{x=-W} \right\}^{2/3} \qquad \text{IV-(58)}$$

[where optical waveguide notation has been used for the right-hand side of Eq. (58) instead of Budden's radio ionosphere notation]. Using Eq. (58), Eq. (57) becomes

$$B >> 1.17 \qquad \text{IV-(59)}$$

For optical waveguides, Eq. (58) almost always satisfies Eq. (59). That is, if the linearity criterion given by Eq. (46) is satisfied for TE modes, then it is satisfied for TM modes also, to an extremely good approximation. For instance, if the index profile is exponential such that

$$n_g(x) = n_o + \Delta n \left[1 - e^{d(x+D)} \right] / (1 - e^{dD}) \qquad \text{IV-(60)}$$

then Eq. (58) satisfies Eqs. (1) - (3). If the steep-index-profile values used with Eq. (27) are chosen, and the e^{-1} point of the index profile occurs when $x = -D/2 = -0.5\mu m$, then $d = 2\mu m^{-1}$. Also, letting $\eta = n_o = 1.5$ and $W = D$, (i.e., the wave is propagating nearly at cutoff),

then $b'(-W) = b'(-D) = -\Delta nd/(1 - e^{dD}) = 0.313$. Thus, from Eq. (56), $B = 51$, which certainly satisfies Eq. (57). For waveguide index profiles with $\Delta n = 0.01$ over a distance $D = 10\mu m$, B has the value of 1089.

It might be mentioned that in the radio ionospheric case, $(\Delta n)^2$ may not be small but may approach n_o^2. In such a situation, the problem is more complicated, and Budden discusses it in some detail.[11] As he explains it, the field component E_x imparts vertical motions to the electrons and if E_x is large, then within one cycle any one electron sees a different E_x at different levels. In other words, the electron does not see a constant force. Hence its motion does not vary harmonically with time and is therefore anharmonic. Because of this, some energy goes into harmonics, and the original wave is attenuated. As a result, the phase factor encountered upon reflection (or total internal refraction) must have an imaginary component to account for this loss in the TM case. Further, the phase change is less than $\pi/2$. Maximum attenuation (i.e., conversion to harmonics) occurs when $B \sim 0.45$, in which case only about 72% of the incident energy is reflected by the ionosphere.

For the optical waveguide case, B is very large; hence a negligible loss should occur per reflection (i.e., per total internal refraction). However, unlike radio waves which reflect once from the ionosphere, light in a waveguide might be reflected many times during its passage within the guide. Perhaps, for certain index profiles, enough energy can be converted to higher harmonics to appreciably attenuate the TM wave. And perhaps, with the proper

phase matching, non-linear effects such as second-harmonic generation or parametric upconversion might take place. As Budden points out,[11] the problem of TM attenuation is very difficult, requiring numerical computations. Since from Eq. (16) or (17) [and Eq. III-(6)], $\Theta(x)$ is very small for small Δn, then, for a propagation length of a few millimeters, the number of total internal refractions will be quite small. Further, since B is so large (on the order of 10^2 or more) for such profiles, a negligible loss per refraction should occur.[11] Thus, no significant attenuation should occur for a TM wave in a slowly varying index optical waveguide which is only a few millimeters in length (such as the zinc sulfide films used in this research project). It may be that waveguides with steeper profiles and longer lengths will produce significant attenuation of a TM wave.

It is now possible to write down the solutions for $\vec{E}_{TM}$ and $\vec{H}_{TM}$. For the region $x \geq 0$, the solutions are given by Eqs. III-(46), C-(224), and C-(227). For the region $(-W + \Delta) \leq x \leq 0$, the solutions are the WKBJ solutions given by Eqs. III-(47), C-(225), and C-(228). For the region $(-W - \Delta) \leq x \leq (-W + \Delta)$, the solutions involve the Airy integral function:

$$H_{oy}(x) = n_g \, i\beta \, K_{TE} \, Ai(\zeta) \qquad \text{IV-(61)}$$

The exact form of E_{ox} and E_{oz} is given in Eqs. D-(151) and D-(152). Finally, for the region $x \leq (-W - \Delta)$, the WKBJ exponentially decaying solution is

$$H_{oy}(x) = -n_g \, |q|^{-1/2} \, A_4 \, e^{-k_o \int_{-W}^{x} |q| \, dx} \qquad \text{IV-(62)}$$

where A_4 is defined in Eq. D-(144). The exact form of F_{ox} and F_{oz} in this region is given in Eqs. D-(154) and D-(155).

From the analysis for TE and TM modes given above and in Appendix D, all the results obtained from the quasi-geometric ray-optics approach can be verified from considerably more rigorous mathematics. Thus it can be concluded that a practical and accurate understanding of the phenomenon of total internal refraction in optical waveguides can be obtained by using physical insight arguments without the necessity of finding the actual solutions for $\vec{E}$ and $\vec{H}$. Of course, the conditions under which such arguments are valid must be determined from the more rigorous approach.

CHAPTER V

GUIDING OF LIGHT IN EPITAXIAL ZINC SULFIDE FILMS

Introduction

Cubic zinc sulfide (also called β-ZnS, zincblende, or, for naturally occurring crystals, zincblende or sphalerite) has a bulk refractive index[16] $n_Z = 2.35$ at $\lambda_o = 0.6328\mu m$. The epitaxial zinc sulfide films fabricated by the Central Research Laboratories of Itek Corporation are grown on gallium arsenide (GaAs) substrates.[17,18] The bulk refractive index of GaAs is complex:[19] $\tilde{n}_G = n_G + i\kappa_G = 3.87 + 0.087i$ at $\lambda_o = 0.6328\mu m$. [Although the imaginary part κ_G of the index is only about 2% of the real part n_G, the absorption coefficient[20] α is approximately $1.7 \times 10^4\ cm^{-1}$, so, for $\lambda_o = 0.6328\mu m$, no light propagation can take place in GaAs crystals of length L greater than a few micrometers.] Since $\kappa_G << n_G$, then $|\tilde{n}_G| \simeq n_G$, and the reflection coefficient[21] R has little dependence on κ_G. Further, since $n_G > n_Z$, then light propagation in an epitaxial zinc sulfide film on a gallium arsenide substrate cannot take place solely by means of total internal reflection. Instead, either partial internal reflection or, possibly (as discussed below), total internal refraction must be the mechanism for guiding light at or near the ZnS-GaAs boundary.

Total Internal Refraction

Cubic GaAs has a lattice spacing of 5.64 Å; cubic ZnS has a natural spacing of 5.41 Å (Ref. 17). If there is a perfect lattice match at

the ZnS-film/GaAs-substrate boundary, then the spacing of the ZnS film lattice will change from 5.64 Å to a more closely packed 5.41 Å, a difference of 4.25%. It seems possible that, because of the greater ZnS lattice spacing at the ZnS/GaAs boundary, a lower ZnS density as well as a lower ZnS refractive index should result. Then, as the epitaxial ZnS lattice changes from 5.64 Å to 5.41 Å, an index gradient could exist within the guide, with a higher value of the index n_Z at distances increasingly farther from the ZnS/GaAs boundary. Thus, light could propagate in the ZnS film by means of total internal refraction at some level above the ZnS/GaAs boundary.

There is a second possible mechanism which might give rise to a positive index gradient for n_Z in addition to the lattice mismatch. The coefficient of linear thermal expansion (at 200°C) of zinc sulfide is 6.3×10^{-6}/°C: for gallium arsenide the value is 6.5×10^{-6}/°C (Ref. 22). After crystal film growth at typically 400°C, the ZnS and GaAs must cool down simultaneously. In doing so, the smaller thermal expansion of the ZnS film causes a small bending of the ZnS/GaAs combination, such that the ZnS film is on the inner diameter. As a result, the density of the ZnS film is perhaps slightly higher at the ZnS/superstrate boundary. Such a bending was observed previously by scientists at Itek (unpublished), and a "focal length" f of approximately one meter was observed for a collimated beam of laser light. (To the eye, the surface appears perfectly flat.) Thus the radius of curvature of the bending is 2f. For a ZnS film thickness of W = 20μm or so (a typical value of the Itek-grown films used for the light guiding experiments), then only a small change in the radius of curvature 2f is present between the top and bottom of the film. For a crystal length dimension L_T at the top surface, then the angle

subtended by the top surface of the crystal is $L_T/4\pi f$: for the bottom surface, the same angle is subtended and can be written as $L_B/2\pi(2f + W)$, where L_B is the length of the film along the bottom of the ZnS film. Equating the above two expressions leads to $L_B = L_T(1 + W/2f)$. Since $W/2f \sim 10^{-5}$, it is reasonable that any index gradient due to a differential thermal expansion should be negligible compared to an index gradient caused by the lattice mismatch.

The question, then, is how far into the ZnS film does the lattice mismatch have any effects. Experiments by other authors[23,24] with zinc sulfide films deposited on glass and fused silica substrates have confirmed the presence of index inhomogeneities extending anywhere from 100 Å to 800 Å (i.e., from 20 to 150 lattice spacings) away from the substrate/film boundary. Of course, since the substrates were amorphous, the zinc and sulfur atoms had complete freedom to align themselves: in fact, the ZnS films were nearly amorphous themselves, showing a minimum of crystalline structure. On the other hand, the epitaxial growth of zinc sulfide films on gallium arsenide substrates requires that the ZnS molecules align themselves according to the cubic-structured GaAs lattice. Hence, the possibility exists that any inhomogeneity due to a lattice mismatch could extend more than a 1000 Å into the ZnS film.

Partial Internal Reflection

Partial internal reflection at the ZnS/GaAs boundary is another possible mechanism for guiding of light in a zinc sulfide thin film. Such a light guide is a leaky one[25] since some energy is lost upon reflection by waves at the ZnS/GaAs boundary. Ulrich and Prettl[25]

discuss leaky light guides in detail and arrive at the following relationships for a given mode number m:

$$\beta_m = k_0 \eta_m - i\alpha_m \quad \text{(TE and TM modes)} \qquad \text{V-(1)}$$

$$\eta_m = n_g - \frac{\pi^2}{2n_g}\frac{(m+1)^2}{(k_0 W_{eq})^2} \quad \text{(TE and TM modes)} \qquad \text{V-(2)}$$

$$\alpha_m^{TE} = \frac{\pi^2 k_0}{n_g(n_s^2 - n_g^2)^{1/2}}\frac{(m+1)^2}{(k_0 W_{eq})^3} \quad \text{(TE modes)} \qquad \text{V-(3)}$$

$$\alpha_m^{TM} = \frac{n_s^2 \pi^2 k_0}{n_g(n_s^2 - n_g^2)^{1/2}}\frac{(m+1)^2}{(k_0 W_{eq})^3} \quad \text{(TM modes)} \qquad \text{V-(4)}$$

where

$$k_0 W_{eq} = k_0 W + \frac{1}{(n_g^2 - n_a^2)^{1/2}} \quad \text{(TE modes)} \qquad \text{V-(5)}$$

$$k_0 W_{eq} = k_0 W + \frac{n_a^2}{n_g^2}\frac{1}{(n_g^2 - n_a^2)^{1/2}} \quad \text{(TM modes)} \qquad \text{V-(6)}$$

Eq. (2) can be considered as a characteristic waveguide equation for leaky light guides. The quantities n_a, n_s, n_g and W are defined in Figure II-1 except that, for leaky guides,

$$n_s > n_g > \eta_m > n_a \qquad \text{V-(7)}$$

The quantity β_m is similar to the wavevector component along the $\hat{z}$-axis [from Eq. II-(22)], only now a provision is made for attenuation of the light by including the term $-i\alpha_m$.

The quantity W_{eq} is the "equivalent thickness"[25] of the guide and is the result of the fact that the totally internally reflected wave at the $n_a - n_g$ boundary penetrates slightly into the n_a region. For λ_o = 0.6328μm and W = 20μm, then $k_o W = 200$; for $n_g = n_Z = 2.35$ and $n_a = 1.63$ (the largest index of the superstrate materials used in waveguiding experiments), then $(n_g^2 - n_a^2)^{-1/2} = 0.6 << k_o W$. Hence, for all practical purposes, W can be used instead of W_{eq} in Eqs. (1) - (6) for both TE and TM modes in Itek-grown epitaxial zinc sulfide light guides.

Eqs. (1) - (4) were derived[25] under the assumption that the imaginary part κ_s of the refractive index $\tilde{n}_s$ is much less than the real part n_s, i.e., $\kappa_s << n_s$ and $|\tilde{n}_s| \sim n_s$. As explained in the introduction to this chapter, the values of n_G and κ_G satisfy this assumption. Using the value $n_s = n_G = 3.87$ together with λ_o = 0.6328μm and W = 20μm, then, from Eqs. (3) and (4)

$$\alpha_m^{TE} = (m+1)^2 \left(1.7 \times 10^{-3}\ mm^{-1}\right) \qquad \text{V-(8)}$$

$$\alpha_m^{TM} = (m+1)^2 \left(4.6 \times 10^{-3}\ mm^{-1}\right) \qquad \text{V-(9)}$$

Since the samples used for light guiding experiments had lengths of a few millimeters, negligible attenuation should occur for the lower-order modes ($m < 5$).

Ignoring the term $i\alpha_m$ in Eq. (1), then from Eq. II-(22), η_m can be written as

$$\eta_m = n_g \sin \theta_m \qquad \text{V-(10)}$$

where θ_m is the value of θ in Figure II-1 for a given mode m. Combining Eqs. (2) and (10),

$$n_g(1-\sin\theta_m) = \frac{\pi^2}{2n_g}\,\frac{(m+1)^2}{(k_o W)^2} \qquad \text{V-(11)}$$

or

$$\sin\theta_m = 1 - \frac{\lambda_o^2}{8n_g^2 W^2}(m+1)^2 \qquad \text{V-(12)}$$

Using the values of λ_o, $n_g = n_Z$, and W used above,

$$\sin\theta_m = 1 - (2.2\times10^{-5})(m+1)^2 \qquad \text{V-(13)}$$

It may be seen that, to a very good approximation, θ_m is very nearly equal to (but always slightly less than) $\pi/2$ -- propagation straight down the guide. This is true for all but the highest order modes, which, from Eqs. (8) and (9), are strongly attenuated anyway.

CHAPTER VI

THE LINEAR ELECTRO-OPTIC EFFECT OF EPITAXIAL ZINC SULFIDE FILMS

The Principal Axes

The optical properties of a crystal can be described in terms of an optical indicatrix, a refractive index ellipsoidal surface. Since zinc sulfide is isotropic, its indicatrix can be represented by the following equation:[26]

$$\frac{x^2+y^2+z^2}{n_o^2} = 1 \qquad \text{VI-(1)}$$

which is the equation of a sphere. Here $n_o = n_Z$, the bulk refractive index. Upon application of an external electric field $\vec{e}$ (not to be confused with the electromagnetic field $\vec{E}$ of a light wave) for the linear electro-optic effect, then the indicatrix becomes[26]

$$\frac{x^2+y^2+z^2}{n_o^2} + 2r_{41}(e_x yz + e_y zx + e_z xy) = 1 \qquad \text{VI-(2)}$$

Here r_{41} is the electro-optic coefficient of zinc sulfide.

The epitaxial films grown by Itek Corporation's Central Research Laboratories are oriented in the (100) direction, defined as the $\hat{x}$ direction in Figure VI-1. On top of this film an electrode is deposited. Thus, the applied field across the film is in the $\hat{x}$ direction only, i.e., $e_y = e_z = 0$. Eq. (2) then reduces to

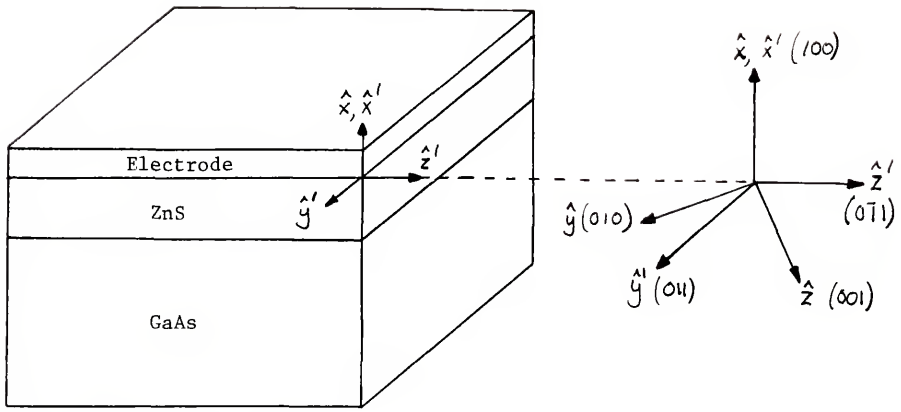


Figure VI-2. Crystal orientation of epitaxial zinc sulfide films grown by the Central Research Laboratories of Itek Corporation. The principal axes are given by the $\hat{x}'$, $\hat{y}'$, and $\hat{z}'$ directions which are also the normals to the natural cleavage planes.

$$\frac{x^2+y^2+z^2}{n_o^2} + 2 r_{41} e_x yz = 1 \qquad \text{VI-(3)}$$

By a proper rotation of the index ellipsoid, a new set of principal axes ($\hat{x}'$, $\hat{y}'$, $\hat{z}'$) can be obtained such that

$$\frac{x'^2}{n_x'^2} + \frac{y'^2}{n_y'^2} + \frac{z'^2}{n_z'^2} = 1 \qquad \text{VI-(4)}$$

Comparing Eqs. (3) and (4) it may be seen that the $\hat{x}$ axis remains unchanged, i.e.,

$$\hat{x}' = \hat{x} \qquad \text{VI-(5)}$$

and

$$n_x' = n_o \qquad \text{VI-(6)}$$

The problem thus reduces to calculating the angle Ω through which the $\hat{y}$ and $\hat{z}$ axes must be rotated to $\hat{y}'$ and $\hat{z}'$. As shown in Appendix E, $\Omega = \pi/4$. The new axes $\hat{x}'$, $\hat{y}'$, and $\hat{z}'$ are shown in Figure VI-1. They coincide with the normals to the natural cleavage planes of cubic zinc sulfide.

In the course of obtaining the principal axes, the dependence of the principal refractive indexes n_x', n_y', and n_z' on the applied field e_x are also obtained in Appendix E. The results to first order are

$$n_x' = n_o \qquad \text{VI-(7)}$$

$$n_y' = n_o - n_o^3 r_{41} e_x/2 \qquad \text{VI-(8)}$$

$$n_z' = n_o + n_o^3 r_{41} e_x/2 \qquad \text{VI-(9)}$$

where $n_o^3 r_{41} e_x/2 \ll n_o$. It may be seen that n_y' and n_z' have a linear dependence on the applied field e_x.[26]

The Transverse Electro-Optic Effect

For propagation in either the $\hat{z}$ or $\hat{z}'$ direction [the (001) and (0$\bar{1}$1) directions, respectively], the light has a $\vec{k}$ - vector component $\beta = k_o n_o \sin\theta$ which is perpendicular (i.e., transverse) to the applied electric field $\vec{e}$. This results in the transverse electro-optic effect. [This is not to be confused with the transverse electric (TE) mode of propagation.] The light also has a $\vec{k}$-vector component $k_o n_o \cos\theta$ which is parallel to $\vec{e}$, resulting in the longitudinal electro-optic effect.

In order to get intensity modulation (as opposed to pure phase modulation) when the ZnS guide is placed between crossed polarizers, it is necessary that the polarization of the incident light be altered. In order to determine if changes in polarization will occur, all incident polarization components must first be resolved along the principal axes (so long as the axes are perpendicular to the net direction of light propagation). If there are two polarization components each parallel to a different principal axis which is perpendicular to the net direction of travel, then each polarization component will "see" a different refractive index if an electric field is applied to the crystal. As a result, each component will have a different phase retardation, and the light will become elliptically polarized as it propagates down the guide. The net effect is to rotate the plane of polarization of the incident light.

A combination transverse and longitudinal electro-optic effect can in general be very complicated. However, for optical waveguides, there are several factors which tend to reduce the complexity.

First of all, the angle Θ will be very nearly 90° for epitaxial zinc sulfide films which guide light at or near the ZnS/GaAs boundary either by partial internal reflection or by total internal refraction. In the former case, this is obvious from Eq. V-(13). Since the longitudinal component depends on the cosine of Θ, it will be negligible for $\Theta \sim \pi/2$. In the latter case, the change in index due to a lattice mismatch is probably no more than 4.25%, i.e., $\Delta n = 0.0425 n_Z$. From Eqs. IV-(16) and D-(73), $\sin\Theta(0)$ cannot be any less than $n_Z/(n_Z + \Delta n) = 0.96$, i.e., $\Theta(0)$ cannot be any less than 74°. Of course, $\Theta(x)$ changes from $\Theta(0)$ to $\Theta(-W) = 90°$. Hence, very roughly, the minimum "average" value of $\Theta(x)$ would be 82°. Since $\sin 82° = 0.99$ and $\cos 82° = 0.14$, it may be seen that the longitudinal component (from cos 82°) is, at the very most, only 14% of the transverse component.

Even more significantly, the voltage required to produce a longitudinal electro-optic effect in a thin-film waveguide is orders of magnitude greater than the voltage necessary to produce a transverse effect. For a pure longitudinal effect, the half-wave voltage (defined as the applied voltage required to change the plane of polarization of the incident light by 90°) is given by[26]

$$(V_{\lambda/2})_{Long} = \frac{\lambda_o}{2 n_o^3 r_{41}} \qquad \text{VI-(10)}$$

For a pure transverse effect, the half-wave voltage is[26]

$$(V_{\lambda/2})_{Trans} = \frac{\lambda_o}{n_o^3 r_{41}} \frac{W}{L} \qquad \text{VI-(11)}$$

$$= \frac{2W}{L} (V_{\lambda/2})_{Long} \qquad \text{VI-(12)}$$

Using λ_o = 0.6328μm, $n_o = n_Z$ = 2.35, and r_{41} = 2.07 x 10^{-10} cm/volt (Ref. 26), then $(V_{\lambda/2})_{Long}$ = 1.2 x 10^4 volts. For W = 20μm and L = 1cm, then $(V_{\lambda/2})_{Trans} = 10^{-3} (V_{\lambda/2})_{Long}$ = 47 volts. Thus, with an applied voltage of 50 volts or so, a large rotation of the polarization will occur as a result of the transverse electro-optic effect: a negligible effect will be observed for the longitudinal effect.

Further, since optical waveguides are quite thin, dielectric breakdown will occur before a very large voltage can be applied to the thin-film material. For zinc sulfide, the breakdown voltage for practical purposes is approximately 10^5 volt/cm = 10 volt/μm (Ref. 27). Thus, a 20μm-thick zinc sulfide film would break down long before any significant longitudinal effect occurred.

Thus it can be concluded that the transverse electro-optic effect is the primary mechanism for electro-optic modulation in zinc sulfide optical waveguides. It might be pointed out that, since $\beta = k_o n_Z \sin\Theta$ is a constant for homogeneous as well as inhomogeneous waveguides (as explained in Chapters II-IV), the transverse component of the $\vec{k}$-vector is thus a constant for either case.

Intensity Modulation for Propagation in the $(0\bar{1}1)$ Direction

1. TE Modes

For a pure TE mode, the electric vector $\vec{E}$ is polarized parallel to the $\hat{y}'$ axis, i.e.,

$$\vec{E}_{TE} = \hat{y}' E_{y}' \qquad \text{VI-(13)}$$

$$= \hat{y}' E \qquad \text{VI-(14)}$$

where $E = E_o(x)\, e^{i(\omega t - \beta z)}$. The wave will "see" an index $n_y' = n_o = n_Z$ with zero applied field e_x, and an index $n_y' = n_Z - n_Z^3\, r_{41} e_x/2$ with an applied field. Thus, phase modulation can occur for the TE case. However, since the light is polarized parallel to only one principal axis, then for any value of e_x the $\vec{E}$-field sees only one index of refraction at any one time. Hence the polarization will not be altered, and no intensity modulation is possible.

2. TM Modes

For a pure TM mode, the electric vector $\vec{E}$ can be written as

$$\vec{E}_{TM} = \hat{x}' E_x' + \hat{z}' E_z' \qquad \text{VI-(15)}$$

$$= \hat{x}' E \sin\theta_{TM} - \hat{z}' E \cos\theta_{TM} \qquad \text{VI-(16)}$$

Since the propagation is in the $\hat{z}'$ direction, it may be seen that there is only one polarization component along a principal axis perpendicular to the net direction of travel. Hence, the polarization will not be altered, so no intensity modulation is possible; phase modulation also is not possible.

3. TE + TM Modes

For TE + TM incident light, the $\vec{E}$-vector can be written as

$$\vec{E}_{TE+TM} = \vec{E}_{TE} + \vec{E}_{TM} \qquad \text{VI-(17)}$$

$$= (\hat{x}' E_x' + \hat{y}' E_y' + \hat{z}' E_z') \qquad \text{VI-(18)}$$

$$= \hat{x}'(E/\sqrt{2}) \sin\theta_{TM} + \hat{y}'(E/\sqrt{2}) - \hat{z}'(E/\sqrt{2}) \cos\theta_{TM} \qquad \text{VI-(19)}$$

In this situation, with an applied field e_x, the speed of propagation for E_x' is different from the speed of propagation for E_y'. Thus after travelling a length L through the crystal, unequal phase shifts δ_x' and δ_y' are introduced:

$$E_x' = (E_0 \sin\theta_{TM}) \, e^{i(\omega t - \delta_x')} \qquad \text{VI-(20)}$$

$$E_y' = E_0 \, e^{i(\omega t - \delta_y')} \qquad \text{VI-(21)}$$

where

$$\delta_x' = \beta_{TM} L \qquad \text{VI-(22)}$$

$$\delta_y' = \beta_{TE} L \qquad \text{VI-(23)}$$

and

$$\beta_{TM} = k_0 n_x' \sin\theta_{TM} \qquad \text{VI-(24)}$$

$$\beta_{TE} = k_0 n_y' \sin\theta_{TE} \qquad \text{VI-(25)}$$

The values of θ_{TE} and θ_{TM} are to be determined from the appropriate characteristic waveguide equation, depending on the mechanism for light guiding.

It is now assumed that the zinc sulfide light guide is placed between crossed polarizers (where the input polarizer is oriented so as to pass a TE + TM wave). Upon application of a field e_x, the E_x' and E_y' components, each at angle 45° to the analyzer axis, have components $-E_x'/\sqrt{2}$ and $+E_x'/\sqrt{2}$ along the analyzer axis. The amplitude E_A of the light passed by the analyzer is thus

$$E_A = \left(-E_x' + E_y'\right)/\sqrt{2} \qquad \text{VI-(26)}$$

where the $\hat{z}'$ component has been neglected since $\Theta_{TM} \sim \pi/2$, as explained above. From Eqs. (20) and (21), Eq. (26) becomes

$$E_A = \left(E_o/2\right)\left(-\sin\theta_{TM}\, e^{i\delta_x'} + e^{i\delta_y'}\right)e^{i\omega t} \qquad \text{VI-(27)}$$

The mean (i.e., time-averaged) intensity I transmitted by the analyzer is

$$I = E_A E_A^*/2 \qquad \text{VI-(28)}$$

From Eq. (27) this results in

$$I = \left(E_o^2/8\right)\left\{\sin^2\theta_{TM} + 1 - \sin\theta_{TM}\left[e^{i\left(\delta_x' - \delta_y'\right)} + e^{-i\left(\delta_x' - \delta_y'\right)}\right]\right\} \qquad \text{VI-(29)}$$

Defining

$$I_o \equiv E_o^2/2 \qquad \text{VI-(30)}$$

then

$$I = (I_o/4)\left[\sin^2\theta_{TM} + 1 - 2\sin\theta_{TM}\cos(\delta_x' - \delta_y')\right] \qquad \text{VI-(31)}$$

Using Eqs. (22) - (25) together with Eqs. (7) and (8), then the phase shifts can be written as

$$\delta_x' = k_o n_o L \sin\theta_{TM} \qquad \text{VI-(32)}$$

$$\delta_y' = k_o (n_o - n_o^3 r_{41} e_x/2) L \sin\theta_{TE} \qquad \text{VI-(33)}$$

Thus,

$$\delta_x' - \delta_y' = \delta_o + k_o n_o^3 r_{41} L e_x/2 \qquad \text{VI-(34)}$$

where

$$\delta_o \equiv k_o n_o L (\sin\theta_{TM} - \sin\theta_{TE}) \qquad \text{VI-(35)}$$

Further, since

$$e_x = V/W \qquad \text{VI-(36)}$$

where V is the voltage across the crystal, then

$$\delta_x' - \delta_y' = \delta_o + k_o n_o^3 r_{41} VL/2W \qquad \text{VI-(37)}$$

$$= \delta_o + \pi n_o^3 r_{41} VL/\lambda_o W \qquad \text{VI-(38)}$$

If the approximation is made that

$$\theta_{TM} \approx \theta_{TE} \approx \pi/2 \qquad \text{VI-(39)}$$

which is valid for epitaxial zinc sulfide light guides (as explained above), then Eqs. (31) and (38) become

$$I = I_0 \sin^2 \left(\delta_x' - \delta_y' \right) \qquad \text{VI-(40)}$$

and

$$\left(\delta_x' - \delta_y' \right) = \pi n_0^3 r_{41} V L / \lambda_0 W \qquad \text{VI-(41)}$$

This is the same result obtained for the bulk transverse electro-optic effect,[26] and intensity modulation will thus occur as the applied voltage is varied from zero to $(V_{\lambda/2})_{Trans}$.

It should be mentioned that if there is a dielectric layer isolating the metal electrode from the zinc sulfide film, then the voltage V in Eq. (41) will not be the total voltage V_T, but only the voltage across the ZnS film. The relationship between V and V_T can be obtained from elementary physics since the layer and ZnS film are effectively two parallel-plate capacitors in series. The result is

$$V = \frac{W \epsilon_d}{W \epsilon_d + W_d \epsilon_Z} V_T \qquad \text{VI-(42)}$$

where ε_d and W_d are the low-frequency dielectric constant and thickness, respectively, of the layer; ε_Z is the low-frequency dielectric constant of zinc sulfide and has the value 8.37 ± 0.8 picofarad/meter (Ref. 28).

Intensity Modulation for Propagation in the (001) Direction

1. TE Modes

For a pure TE mode, the electric vector $\vec{E}$ is polarized to the $\hat{y}$ axis, i.e.,

$$\vec{E}_{TE} = \hat{y} E_y \qquad \text{VI-(43)}$$

$$= \hat{y}' E_y' + \hat{z}' E_z' \qquad \text{VI-(44)}$$

$$= \left(\hat{y}' - \hat{z}'\right) E/\sqrt{2} \qquad \text{VI-(45)}$$

The polarization components of the incident light must always be at right angles to the direction of propagation. Since the $\hat{y}'$ and $\hat{z}'$ axes are <u>not</u> perpendicular to the net direction of propagation for the $\hat{z}$-propagation case, then the method of resolving the incident polarization along principal axes is not a proper way of examining the problem. Instead, the effective contributions of the axes can be evaluated by resolving these axes along the axis formed by the incident polarization. In this situation, any change $\Delta n_y'$ due to the $\hat{y}'$ axis is identically compensated by any change $\Delta n_z' = -\Delta n_y'$ [from Eqs. E-(17) and E-(18)] due to the $\hat{z}'$ axis. Hence the TE wave "sees" a constant index $n_o = n_Z$. The net effect is no change at all (not even phase modulation when a field is applied to the crystal). In other words, for propagation in the $\hat{z}$ direction, at least for TE modes, the zinc sulfide appears isotropic, with or without an applied electric field.

2. TM Modes

For a pure TM mode, the electric vector $\vec{E}$ can be written as

$$\vec{E}_{TM} = \hat{x} E_x + \hat{z} E_z \qquad \text{VI-(46)}$$

$$= \hat{x}' E_x' + \hat{y}' E_y' + \hat{z}' E_z' \qquad \text{VI-(47)}$$

$$= \hat{x}' E \sin\theta_{TM} - (\hat{y}' + \hat{z}')(E/\sqrt{2})\cos\theta_{TM} \qquad \text{VI-(48)}$$

Since $\theta_{TM} \sim \pi/2$, then Eq. (48) reduces to

$$\vec{E}_{TM} = \hat{x}' E \qquad \text{VI-(49)}$$

which is the same as Eq. (16). Thus no intensity modulation is possible (and, since $n_x' = n_o$ is independent of the applied field e_x, no phase modulation is possible either). Even if θ_{TM} is not quite equal to $\pi/2$, there will still be no net phase change from the $\hat{y}'$ and $\hat{z}'$ terms for the same reasons given for TE modes propagating in the $\hat{z}$ direction.

3. TE + TM Modes

For TE + TM incident light, the $\vec{E}$-vector can be written as

$$\vec{E}_{TE+TM} = \vec{E}_{TE} + \vec{E}_{TM} \qquad \text{VI-(50)}$$

$$= \hat{x}' E_x' + \hat{y}' E_y' + \hat{z}' E_z' \qquad \text{VI-(51)}$$

$$= \hat{x}'(E/\sqrt{2})\sin\theta_{TM} + \hat{y}'(E/2)(1-\cos\theta_{TM}) - \hat{z}'(E/2)(1+\cos\theta_{TM}) \qquad \text{VI-(52)}$$

Again, since $\theta_{TM} \sim \pi/2$, then

$$\vec{E}_{TE+TM} = \hat{x}' E/\sqrt{2} + (\hat{y}' - \hat{z}') E/2 \qquad \text{VI-(53)}$$

Again, the same arguments used for TE and for TM modes lead to the conclusion that the $\hat{y}'$ and $\hat{z}'$ term will produce no net effect. Thus, even though from the appearance of Eq. (52) it might seem that there should be some modulation (either intensity or phase),such is not the case. Hence the $\hat{z}$ direction is an optic axis of cubic zinc sulfide since the refractive index is the same for TE, TM, or TE + TM polarizations.

CHAPTER VII

EXPERIMENTAL PROCEDURE

Sample Preparation

Most of the Itek-grown samples have a circularly shaped zinc sulfide film which does not extend all the way to the edge of the gallium arsenide substrates. Thus, in order to propagate light in the $\hat{z}'$ direction [the $(0\bar{1}1)$ direction shown in Figure VI-1], it is necessary to cleave each sample twice (indicated in Figure VII-1), thereby exposing one edge for input coupling of the light and another edge at the other end for the output. This is accomplished by placing the sharp edge of a one-sided razor blade in the proper orientation on the GaAs side of the ZnS-on-GaAs sample and then tapping the blade slightly. In this way a microscopically smooth edge of the ZnS film is obtained.

For propagation in the $\hat{z}$ direction [the (001) direction], it is necessary to slice the input and output ends of the ZnS-on-GaAs samples (using a diamond saw) at a 45^{o} angle to the normals to the cleavage planes. The sample is then sandwiched between two glass plates and subsequently fine polished. The glass-plate sandwich is used because the first polishing attempts without the plates resulted in a bevelled ZnS edge. As a result, light incident on the film edge was refracted into the GaAs substrate, and propagation down the length of the thin-film guide could not be achieved. Schematic diagrams of the resulting samples for $\hat{z}$ propagation are shown in Figure VII-2.

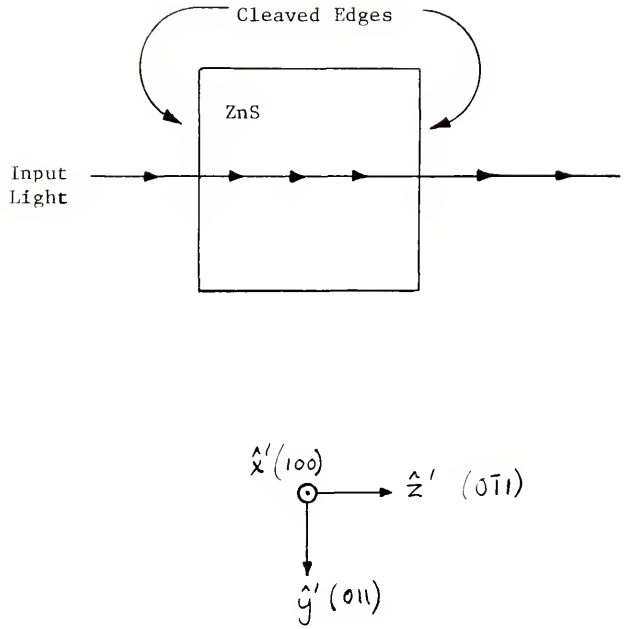


Figure VII-1. Top view of cleaved-edge sample for light propagation in the $\hat{z}'$ direction [the $(0\bar{1}1)$ direction].

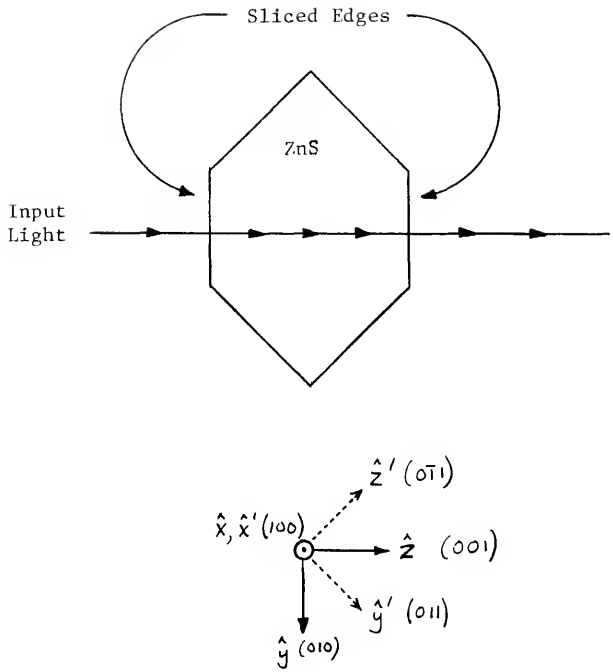


Figure VII-2. Top view of sliced-edge sample for light propagation in the $\hat{z}$ direction [the (001) direction].

For electrode preparation, the top surface of the sample is masked off with Scotch tape, leaving exposed the desired area for the electrode. Then electrically conductive silver paint (Epoxy Products Co., New Haven, Conn.) is applied to this area and allowed to dry, after which the tape is removed. The sample is then mounted as shown in Figure VII-3. A glass substrate is used as an electrically insulating mechanical base. Half of the top surface of the plate is covered with electrically conducting silver paint, and the GaAs side of the ZnS-on-GaAs sample is placed on one portion of this paint before it dries. (Gallium arsenide has a much lower resistivity than zinc sulfide, and therefore it acts as the bottom electrode to the ZnS light modulator.) To the other portion of this paint an electrical lead is affixed with Dupont Duco cement (since the silver paint does not provide much mechanical strength).

On the other half of the glass substrate more silver paint is deposited (electrically isolated from the first half). Then an electrical lead is affixed to this paint and to the top electrode on the ZnS film using some more paint. Finally, another electrical lead is bonded with Duco cement to the silver paint on the glass base. The sample is thus mechanically mounted for waveguiding experiments and electrically connected for electro-optic measurements.

Experimental Set-up

A helium-neon laser is used as the primary light source for the experiments. The output beam of light (λ_o = 0.6328μm) is linearly polarized so that use of a half-wave plate and an angularly calibrated linear polarizer allows easy selection of TE, TM, or TE + TM light. The light is coupled into the guide by means of a 20X microscope objective.

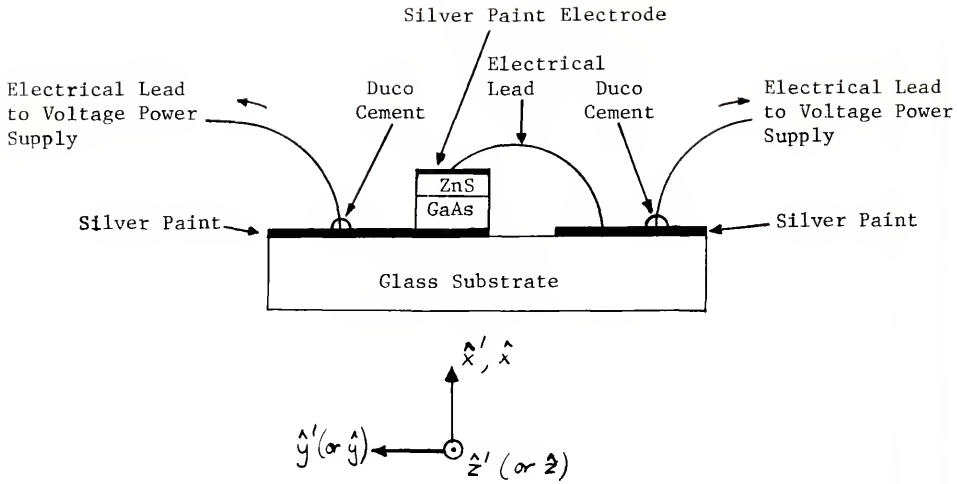


Figure VII-3. End view of mechanical and electrical mounting for waveguiding and electro-optic investigation of zinc sulfide thin-film samples.

At the output end, a microscope, consisting of a 20X or 44X objective and a 10X eyepiece, is used. This experimental set-up is shown schematically in Figure VII-4. The input objective, the ZnS-on-GaAs sample mount, and the output microscope are mounted independently on translational stages each capable of precision motion in three dimensions. The sample mount has an angular degree of freedom also. For the electro-optic modulation experiments an analyzer (with transmission axis 90° to the input polarizer axis) is placed between the ZnS sample and the output microscope.

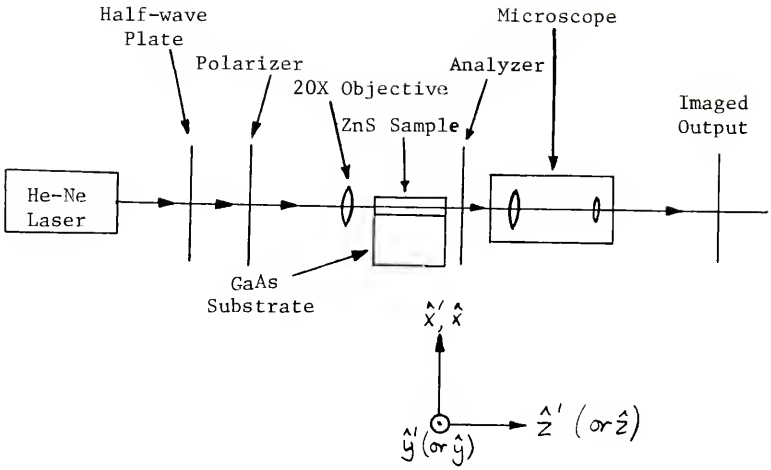


Figure VII-4. Side view of experimental set-up for waveguiding experiments (not drawn to scale).

CHAPTER VIII

EXPERIMENTAL RESULTS AND OBSERVATIONS

Light Guiding

Experiments with light guiding in epitaxial zinc sulfide films were performed to first determine whether partial internal reflection or total internal refraction was the primary mechanism for light guiding at or near the ZnS/GaAs boundary. Preliminary experiments had resulted in an output intensity pattern with a maximum nearer the ZnS/superstrate boundary, suggesting, perhaps, total internal refraction. This is shown in Figure VIII-2: Figure VIII-1 is a photograph of the entire guide illuminated in transmission by white light. Subsequent experiments with several different samples showed that the maximum of the distribution could be shifted to the center of the guide (Figures VIII-3 and VIII-4), and to the bottom of the guide (Figure VIII-5), all for the same mode. (This is essentially an m = 0 mode, although the beginning of a conversion to an m = 1 mode can be seen, especially in Figure VIII-2). This result was obtained by translating the ZnS sample, located at the focal point of the input 20X objective, in the $\hat{x}$ direction (i.e., perpendicular to the direction of light propagation). In this way the effective angle of incidence of the light (defined by the central ray of the cone of light from the objective) could be changed slightly, but not so much as to convert the waveguided light to a higher order mode. Since the entire guide (and not just the top of the guide) could be selectively

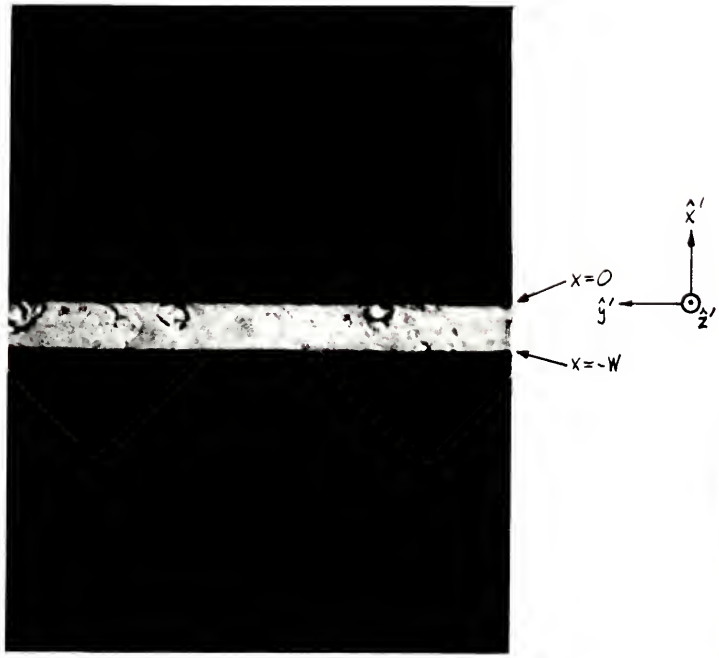


Figure VIII-1. End-view photograph of the cleaved edge of an 18μm-thick zinc sulfide light guide illuminated in transmission by white light. Magnification: 440X. Some microcrystalline defects are evident. The gallium arsenide substrate is below the guide (i.e., below x = -W), the air superstrate above the guide (i.e., above x = 0).



Figure VIII-2. End-view photograph of the output intensity distribution of the cleaved-edge 18μm-thick ZnS guide in Figure VIII-1, but with a higher magnification (780X); wavelength λ_o = 0.6328μm. Figure VIII-2 is a mirror image of Figure VIII-1, as may be seen by comparing the asymmetry of the crystal defect in the center of Figure VIII-1 with the (same) defect in the upper right of Figure VIII-1. It may be noted that the intensity distribution has a maximum near the top of the guide. The mode is m = 0 but approaching m = 1.

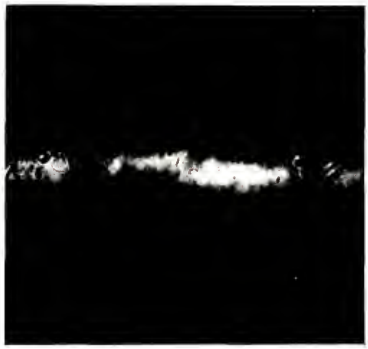


Figure VIII-3. Same as in Figure VIII-2, only now the intensity distribution has a maximum near the center of the guide.

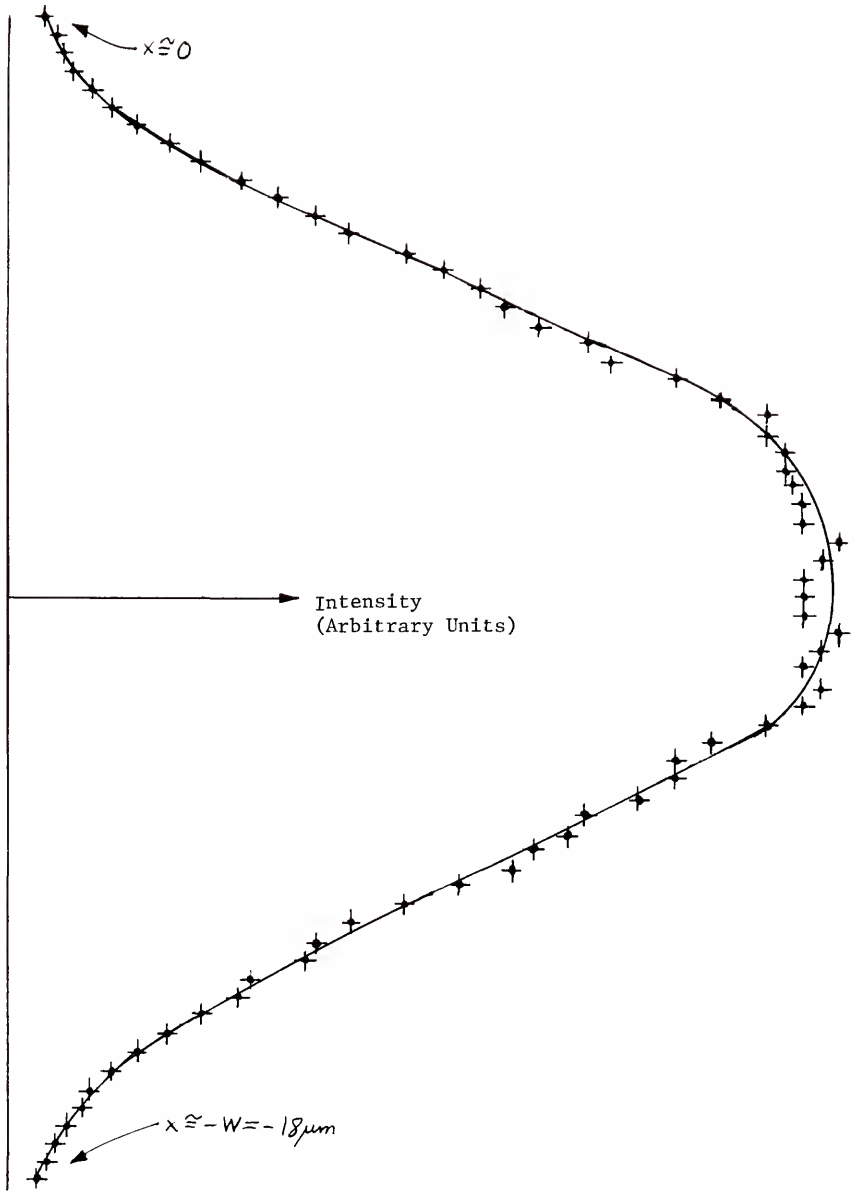


Figure VIII-4. Plot of the intensity distribution in Figure VIII-3 showing the symmetry.



Figure VIII-5. Same as in Figure VIII-2, only now the intensity distribution has a maximum at the bottom of the guide.

"illuminated" in this manner, this suggested that the ZnS films were fairly homogeneous throughout; if some inhomogeneity is present, then the index profile is an extremely slowly varying one. Hence, continual internal refraction is probably not the primary mechanism for light guiding within a ZnS film.

The question remains: what about the effects of the lattice mismatch between the ZnS and the GaAs? During further experiments with various samples illuminated in transmission with white light, an interesting phenomenon was discovered. Inadvertently, the white light bulb used for illumination overfilled the input 20X objective with the result that the ZnS-on-GaAs sample was illuminated in reflection also, i.e., light reflected off the face of the output microscope objective. As a result, it was observed that these samples showed a thin black line at the ZnS/GaAs boundary. This line was not visible (or, at most, was barely discernible) when the sample was viewed in reflected light only. This black line is shown in Figures VIII-6 and VIII-7 and is seen to be about 1μm thick. The thickness of this line did not increase when a voltage of two hundred volts or more was applied to the crystal, as might have occurred if the line was due to some semiconductor junction effect. Thus, it seems reasonable to conclude that this line is a permanent "built-in" part of the ZnS-GaAs structure.

This black line must be either zinc sulfide or gallium arsenide since no other material was present when the ZnS film was epitaxially grown. (If there were some other material, such as oxygen, then a zinc oxide layer would have prevented growth of a good, cubic-structured ZnS film.) If the line were GaAs, then light would strongly reflect from

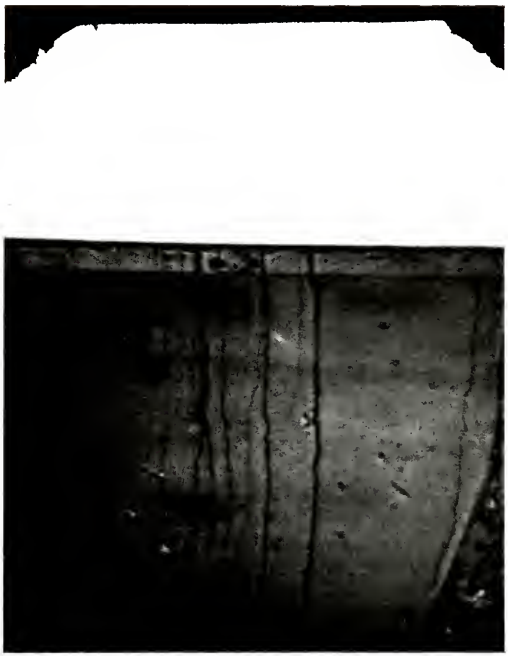


Figure VIII-6. End-view photograph of the cleaved edge of an 11μm-thick zinc sulfide guide on a gallium arsenide substrate illuminated simultaneously in transmission and in reflection with white light. Magnification: 440X. A black line of about 1μm thickness is evident at the ZnS/GaAs boundary. The somewhat thicker black line at the top of the ZnS film is due to a 3μm-thick layer of dielectric parylene used to isolate the metal electrode from the zinc sulfide. Vertical microcrystalline cracks in the sample are also evident.



Figure VIII-7. Same as Figure VIII-6 but with a shorter exposure to enhance the contrast of the black line at the ZnS/GaAs boundary.

it (as seen from the bottom of Figures VIII-6 and VIII-7). This is due to the high index $n_G = 3.87$, resulting in a high normal-incidence GaAs/air reflectivity of 35% or so. Therefore, the black line is of zinc sulfide material. Further, since this black line is prominent only when the white light has travelled the length of the thin film crystal, it seems plausible to also conclude that this line is a result of the lattice mismatch. A 1μm layer corresponds to approximately 2000 lattice spacings. As discussed in Chapter V, it is not unreasonable to expect that the lattice mismatch could have effects extending beyond 1000 Å. Thus, it appears that although an inhomogeneity may exist within an epitaxial zinc sulfide film, it does not extend beyond 1μm from the ZnS/GaAs boundary, a distance of only 5 - 10% of the total film thickness. Thus, 90 - 95% of the film is quite homogeneous.

As stated in the caption to Figure VIII-6, the thicker black line at the top of the ZnS film is due to a 3μm-thick superstrate layer of parylene (a transparent dielectric material obtainable from Union Carbide Corp., Bound Brook, N. J.; refractive index $n_a = 1.63$). Since $n_a < n_Z$, the parylene must be a leaky light guide with losses proportional to the cube of the reciprocal of the guide thickness (cf. Chapter V). Using Eq. V-(3), the attenuation coefficient α_m is approximately 1.3mm^{-1}; since the length of the ZnS crystal and parylene is 5mm, no light can propagate within the parylene but must, instead, eventually leak into the ZnS film. Further, since the parylene is transparent and since only a small amount of light ($\sim$ 4%) will reflect from the parylene/air boundary, the illuminating light reflecting from the output microscope objective must be transmitted in the $-\hat{z}$ direction into the parylene. This is why the parylene appears black.

The same line of reasoning can be applied to the zinc sulfide black line. In addition, the black appearance cannot be attributed to scattering, since forward and backward scattering would tend to illuminate the region as seen through the microscope. Hence, depending on whether or not the inhomogeneous region has a smooth transition to the more homogeneous region, it must be concluded that the inhomogeneity is great enough that either:

1. A very thin and very leaky light guide has been formed (since both the substrate index n_G and the superstrate index n_Z of the rest of the ZnS film exceed the index of the inhomogeneous region), in which case light from the main ZnS guide is guided by total internal reflection from this region: or,
2. Total internal refraction of light from the main ZnS guide into and out of this region is taking place.

For either situation, there is thus a good physical reason, backed up by strong experimental evidence, for believing that a dielectric, inhomogeneous, ZnS isolation layer exists between the main zinc sulfide guide and the gallium arsenide substrate. Thus the main ZnS guide is probably <u>not</u> a leaky light guide, but instead is a true dielectric waveguide. This conclusion contradicts the statements by Martin[29] and by Hall and Yeh[30] based on their results and analyses with epitaxial ZnS samples grown at Itek Corporation and supplied to them for examination. This is not surprising in view of the fact that the inhomogeneous region is on the order of 1μm thick and only the combination of transmission and reflection illumination revealed its existence.

Photographs of mode structures (other than m = 0) for light propagating in the $\hat{z}'$ direction [the $(0\bar{1}1)$direction] are given in Figures VIII-8 through VIII-10, and, for light propagating in the $\hat{z}$ direction

[the (001) direction], in Figures VIII-11 through VIII-13. The various modes were obtained by translating the sample in the $\hat{x}$ direction. It may be seen in Figure VIII-14 that the guide output edge no longer appears rectangular. This is due to the great difficulty in obtaining a well polished edge down to the dimensions of micrometers (which is an entirely different problem from fine polishing a broad surface to such dimensions). As a result, the light is either scattered or refracted out of the field of view of the microscope by the sliced edge. White-light photographs of the sliced edge are not given here since very little of the edge was in focus for a given microscope position. As mentioned in the caption to Figure VIII-11, a pure m = 0 mode could not be obtained for light propagation in the $\hat{z}$ direction.

Finally, it should be mentioned that the mode patterns shown in the figures are representative of TE, TM, or TE + TM modes. It was observed that when the input polarization was changed (e.g., TE to TM), it was necessary to slightly re-adjust the position of the sample with respect to the focal point of the input 20X objective in order to re-obtain the same mode structure.

Electro-Optic Modulation

After becoming familiar with the light guiding properties of epitaxial zinc sulfide films, electro-optic experiments were begun. Initially a large amount (a few milliwatts) of laser light was coupled into the guides. With the samples placed between crossed polarizers and a voltage applied, a transverse electro-optic effect was observed. However, for those samples with parylene between the electrode and the ZnS film, the light transmitted upon initial application of the voltage gradually decreased in intensity with time (over a few seconds). Upon re-application of the voltage (i.e., off then on again), the transmitted



Figure VIII-8. End-view photograph of the cleaved edge of an 18μm-thick zinc sulfide film showing an m = 1 mode pattern (i.e., one horizontal node); wavelength λ_o = 0.6328μm. Magnification: 440X.



Figure VIII-9. Same as Figure VIII-8 but for an m = 2 mode pattern (i.e., two horizontal nodes). On the left an m = 3 mode pattern can be seen.



Figure VIII-10. Same as Figure VIII-8 but for an m = 5 mode pattern.



Figure VIII-11. End-view photograph of the sliced edge of a 20μm-thick zinc sulfide film showing an m = 1 mode pattern; wavelength λ_o = 0.6328μm. Magnification: 440X. (A pure m = 0 mode pattern was not obtainable.)



Figure VIII-12. Same as Figure VIII-11 but for an m=2 mode pattern.



Figure VIII-13. Same as Figure VIII-11 but showing how the edge of a sliced-edge ZnS sample does not appear rectangular because of the difficulty in polishing the edge, a problem which is not present with the cleaved-edge samples.

light reached its previous peak, but again gradually decreased. This is the result of a small photoconductive sensitivity[27,31] in the red region of the visible spectrum due to absorption from small amounts of impurities in the zinc sulfide film. [Since parylene (deposited on top of the ZnS) has an extremely high resistivity, the photoproduced carriers built up at the ZnS/parylene interface reduced the voltage in the ZnS film.] Subsequently, the intensity of the input light was reduced to a few microwatts, after which no further significant photoconductive effect was present to interfere with the electro-optic measurements. (Although the voltages applied were dc only, at modulation frequencies above 1 Hz or so the photoconductive effect should not be significant even for milliwatt or greater intensity levels. This is because the charge carriers generated by the light can no longer follow the changes in the applied electric field.)[31]

For quantitative measurements, a highly light-sensitive photomultiplier tube was used as a detector since the photoconductive relaxation effect took place for high intensities,precluding the use of less sensitive detectors. For qualitative visual observations, a greater light intensity was required, but since a switch was used manually to apply and re-apply the voltages, this did not present any serious difficulties.

Three samples were investigated, as indicated in Table VIII-1. With one sample (Sample #1), modes m = 0,2, and 4 were obtainable in pure form. (Apparently, proper phase conditions for propagating a pure m = 1 or 3 mode were not met.) This sample also resulted in the greatest amount (95%) of light extinction. For the other two samples, only a pure m = 0 mode was obtainable; all other mode patterns were highly multi-moded. Typical results of the observed electro-optic

modulation are presented in Figures VIII-14 through VIII-18 and are summarized in Table VIII-1. Since Sample #2 had a 3μm-thick layer of parylene between the electrode and the ZnS, Eq. VI-(42) was used to calculate the voltage across just the ZnS crystal (using a value of $\varepsilon_d = 3.1$ for the parylene).

From the graphs it may be seen that with the ZnS sample placed between crossed polarizers the minimum is shifted away from the V = 0 situation. This can be attributed to some residual birefringence within the ZnS crystal, as well as to the fact that, as stated in Chapter II (p. 12), a TE + TM wave initially linearly polarized will become elliptically polarized as it propagates down the guide.

For the modulation curves of the m = 2 and 4 mode patterns (Figures VIII-15 and VIII-17), it is evident that some extinction was obtainable. It should be noted, however, that only extinction occurred, i.e., no increased transmission occurred for the $m \neq 0$ modes. Visually, it was observed that the mode pattern would degenerate into a highly multi-moded pattern as the voltage was increased; at the same time, the over-all output intensity first decreased then increased back again to the intensity present for V = 0. Similar measurements for light propagating in the (001) direction gave no modulation of the light for any polarization or any mode, as predicted by theory. However, in the $(0\bar{1}1)$ case, visual observations revealed that for a given mode pattern at V = 0, the pattern was re-established for a voltage approximately halfway between the minimum and the second maximum. This can probably be explained by the fact that the proper phase conditions for propagating a given mode were not met when the voltage corresponding to the maximum output signal (which was multi-mode) was applied. Another way of stating the above is the following. Although some amplitude modulation was

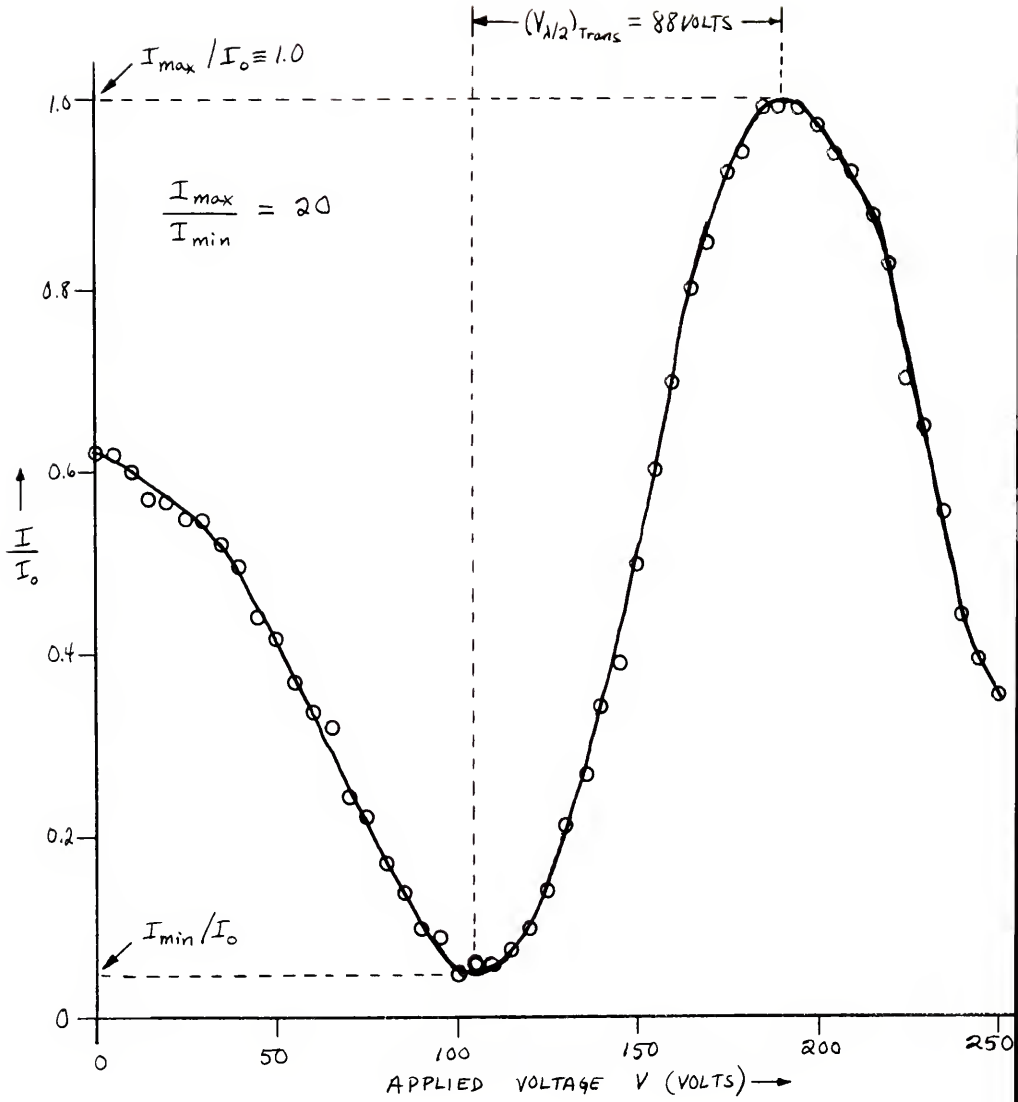


Figure VIII-14. Plot of the transmitted intensity I/I_o versus applied voltage from V = 0 to 250 volts for an m = 0 mode in Sample #1.

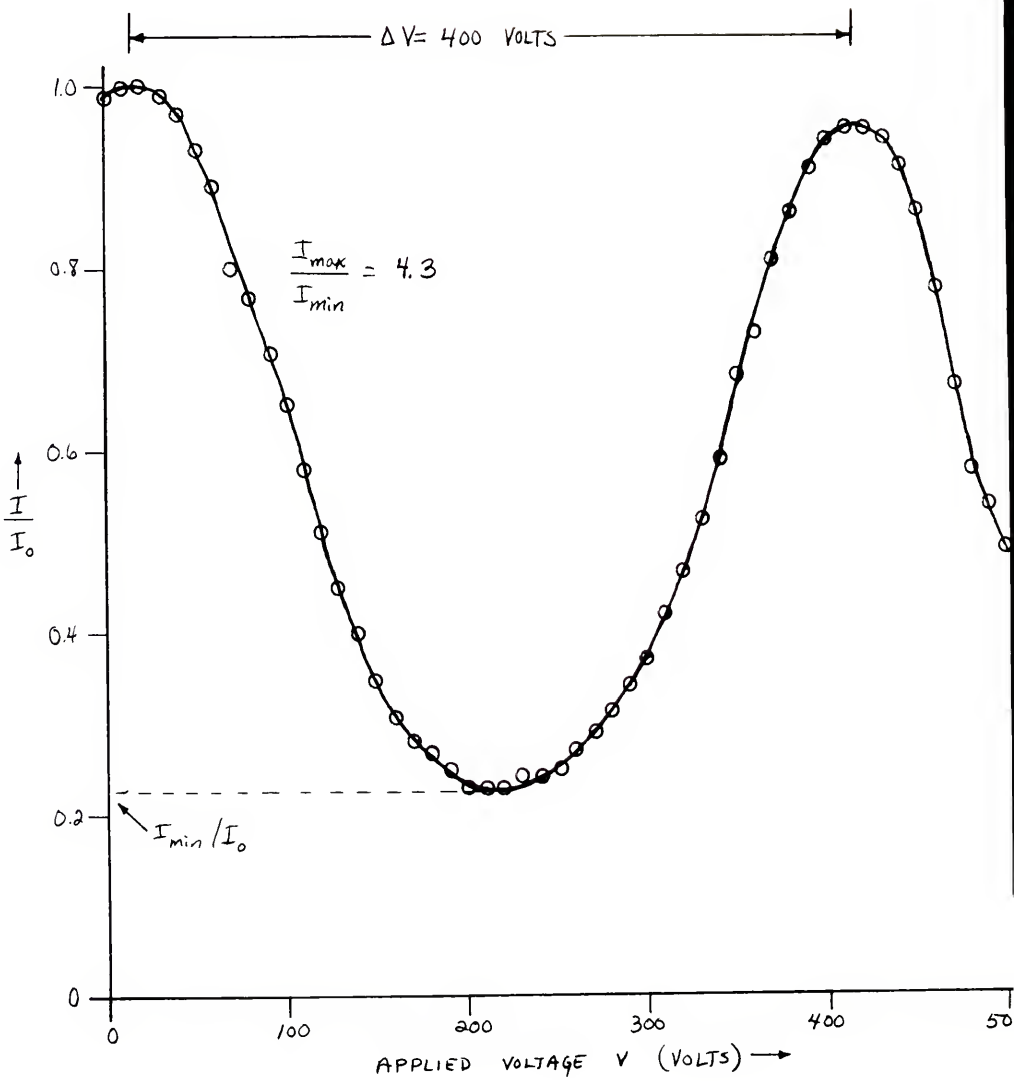


Figure VIII-15. Plot of I/I_o from V = 0 to 500 volts for an m = 2 mode in Sample #1.

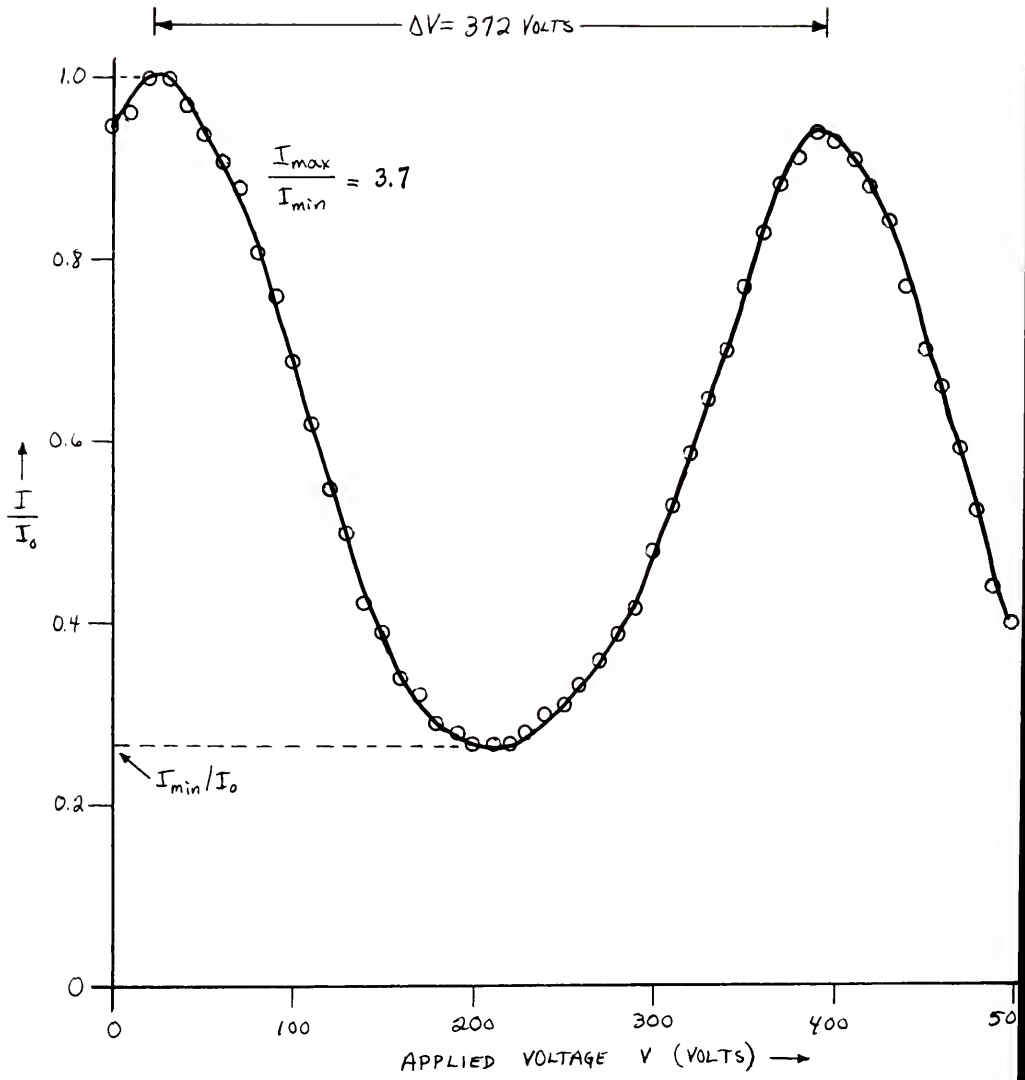


Figure VIII-16. Same as Figure VIII-15 but for an m = 4 mode.

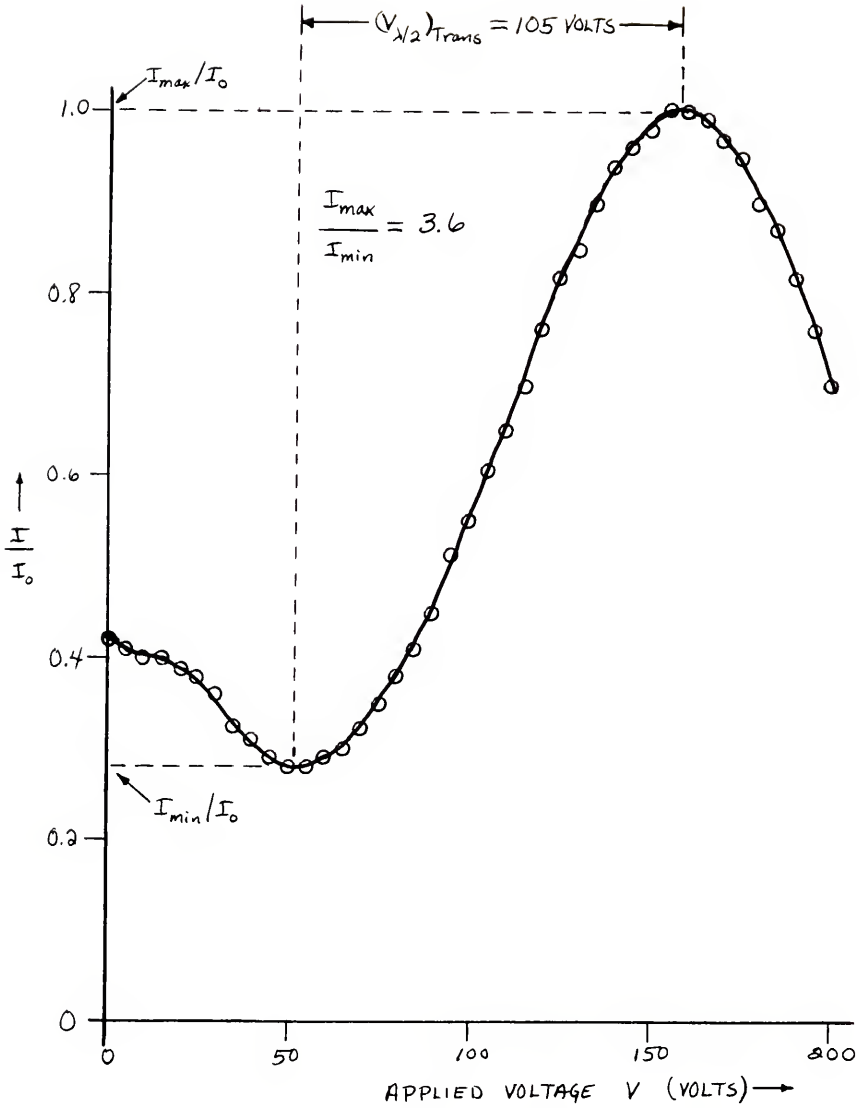


Figure VIII-17. Plot of I/I_o from V = 0 to 200 volts for an m = 0 mode in Sample #2.

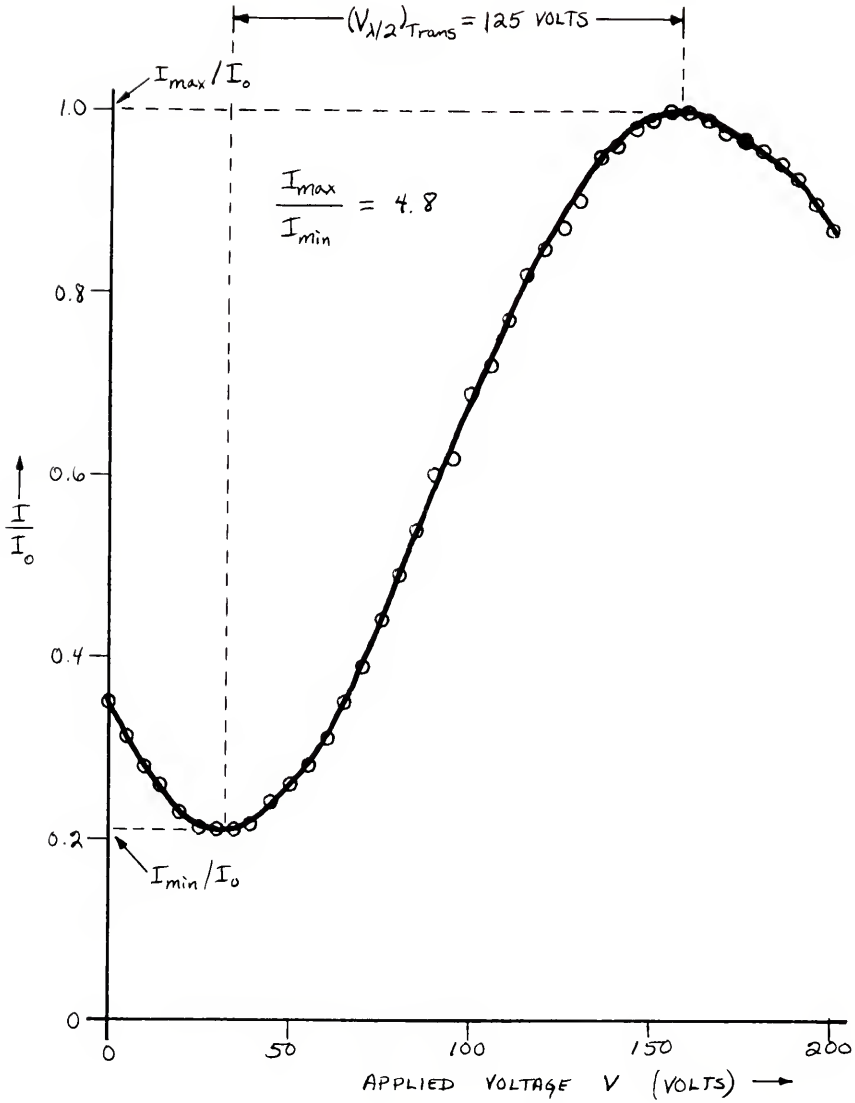


Figure VIII-18. Plot of I/I_o from V = 0 to 200 volts for an m = 0 mode in Sample #3.

Table VIII-1. Summary of Characteristics of Samples Used for Electro-Optic Experiments

Sample #	Thickness W	Interaction Length L	Width h	Experimentally-Determined Value of $(V_{\lambda/2})_{Trans}$	Predicted Value* of $(V_{\lambda/2})_{Trans}$
1	15μm ± 1.5μm	4.0mm ± 0.25mm	3mm	94 volts ± 5 volts	103 volts ± 5 volts
2	25μm ± 1.5μm	8.0mm ± 0.25mm	2mm	80 volts** ± 5 volts	86 volts ± 5 volts
3	17.5μm ± 1.5μm	4.25mm ± 0.25mm	4.5mm	125 volts ± 5 volts	113 volts ± 5 volts

*Obtained by substituting W and L into Eq. VI-(11), where r_{41} = 1.82×10^{-10} cm/volt and n_o = 2.33 (Ref. 18), and λ_o = 0.6328μm.

**The measured voltage V_T was 105 volts ± 5 volts, but since Sample #2 had a 3μm-thick layer of parylene, the value given above was determined from Eq. VI-(42) for the voltage V across the zinc sulfide only.

observed for modes m > 0, the primary effect was phase modulation, resulting in mode conversion. Finally, continued experimentation with the high voltages indicated on the graphs resulted in destructive dielectric breakdown of some of the samples.

In Table VIII-1 it may be noticed that the measured values of the half-wave voltage correspond closely with the values predicted for the bulk-crystal case [from Eq. VI-(11)], which was to be expected as discussed in Chapter VI. It was found that, even though the measured values of $(V_{\lambda/2})_{Trans}$, plus or minus the experimental error, overlapped the predicted values, nevertheless the predicted values were consistently lower than the measured values. Thus, in calculating the expected values, the bulk-crystal value[26] of the electro-optic coefficient r_{41} was not used in Eq. VI-(11). Instead, a measured value[18] of r_{41} for Itek-grown epitaxial zinc sulfide films was used. This value was obtained from longitudinal electro-optic experiments for light incident upon the ZnS from the $\hat{x}$ direction. The result is: $r_{41} = 1.82 \times 10^{-10}$ cm/volt, which is 12% lower than the bulk-crystal value. It may be seen that the results of the transverse electro-optic waveguiding experiments presented here tend to corroborate the measurements of r_{41} obtained previously[18] in separate experiments in which the light was propagated along an entirely different direction in the crystal.

SUMMARY

This dissertation presents the results of a theoretical and an experimental research effort on general index gradient waveguides and their relationship to guiding and electro-optic modulation of light in thin-film zinc sulfide crystals. The basic physics of homogeneous asymmetric waveguides has been reviewed, allowing ready comparison of similar concepts (such as the characteristic waveguide equation) between homogeneous and inhomogeneous optical waveguides. It has been shown that a quasi-geometrical ray-optics approach leads to the same results obtained from more rigorous mathematics, permitting a practical understanding of the phenomenon of total internal refraction and/or reflection in general-index-gradient planar optical waveguides. As a result, the characteristic waveguide equation for describing the propagation of light in such guides can be easily derived. This equation has general applicability to any slowly varying monotonic index profile and allows computation of various waveguide propagation parameters. Of particular note, the phase change of $\pi/2$ (not 0 or π) occurring upon total internal refraction has been explained as being due to a cylindrical focussing of the light within the guide.

Waveguiding and the electro-optic modulation of light in epitaxial zinc sulfide films on gallium arsenide substrates have also been discussed. Two mechanisms for light guiding at the ZnS/GaAs boundary have been described: total internal refraction (resulting from a 4% lattice mismatch) or partial internal reflection (since the refractive index of GaAs exceeds that of ZnS). In the course of waveguiding experiments, the discovery was made of a one-micrometer-thick region of inhomogeneous zinc sulfide which is transparent in visible light and which has a lower

index than the rest of the ZnS. As a result, it is concluded that 1) 90-95% of the zinc sulfide thin film is homogeneous, and 2) due to the thin inhomogeneous layer, the zinc sulfide guide is isolated from the gallium arsenide substrate; therefore the guide is not a leaky one but instead is a true dielectric waveguide. Finally, electro-optic measurements have been presented which agree with those predicted from theory. In addition, results from these experiments have been shown to corroborate previously obtained measurements by other authors on the electro-optic coefficient of Itek-grown epitaxial zinc sulfide films.

APPENDIX A

APPLICATION OF MAXWELL'S EQUATIONS AND VECTOR POTENTIALS

TO HOMOGENEOUS AND INHOMOGENEOUS PLANAR OPTICAL WAVEGUIDES

Maxwell's Equations

Maxwell's equations are:

$$\nabla \cdot \vec{D} = 4\pi\rho \qquad \text{A-(1)}$$

$$\nabla \cdot \vec{B} = 0 \qquad \text{A-(2)}$$

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{B}}{\partial t} \qquad \text{A-(3)}$$

$$\nabla \times \vec{H} = \frac{4\pi\vec{J}}{c} + \frac{1}{c}\frac{\partial \vec{D}}{\partial t} \qquad \text{A-(4)}$$

where $\vec{D} = \varepsilon\vec{E}$, $\vec{B} = \mu\vec{H}$, and c is the speed of light in free space. For a non-permeable medium (μ=1), $\varepsilon = n^2$. For an isotropic dielectric with no charge density (ρ=0) and no currents ($\vec{J}$=0), Eqs. (1)-(4) can be re-written as:

$$\nabla \cdot (n^2\vec{E}) = 0 \qquad \text{A-(5)}$$

$$\nabla \cdot \vec{H} = 0 \qquad \text{A-(6)}$$

$$\nabla \times \vec{E} = -\frac{1}{c}\frac{\partial \vec{H}}{\partial t} \qquad \text{A-(7)}$$

$$\nabla \times \vec{H} = \frac{1}{c} \frac{\partial}{\partial t} (n^2 \vec{E}) \qquad \text{A-(8)}$$

For a homogeneous medium, Eq. (5) may be re-written as:

$$n^2 \nabla \cdot \vec{E} = 0 \qquad \text{A-(9)}$$

Assuming that n is also not a function of time, Eq. (8) may be expressed as:

$$\nabla \times \vec{H} = \frac{n^2}{c} \frac{\partial \vec{E}}{\partial t} \qquad \text{A-(10)}$$

Taking the curl of Eq. (7) and interchanging the curl and time derivative operations gives:

$$\nabla \times \nabla \times \vec{E} = -\frac{1}{c} \frac{\partial}{\partial t} (\nabla \times \vec{H}) \qquad \text{A-(11)}$$

Using a trigonometric identity for the left-hand side (L. H. S.) of Eq. (11), and substituting Eq. (10) into the right-hand side (R. H. S.) yields

$$\nabla (\nabla \cdot \vec{E}) - (\nabla \cdot \nabla) \vec{E} = -\frac{n^2}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \qquad \text{A-(12)}$$

By Eq. (9), the first term of the L. H. S. of Eq. (12) is zero. Writing $\nabla \cdot \nabla$ as ∇^2, Eq. (12) becomes the wave equation:

$$\nabla^2 \vec{E} = \frac{n^2}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} \qquad \text{A-(13)}$$

Similarly, a wave equation for $\vec{H}$ may be obtained by taking the curl of Eq. (10):

$$\nabla \times \nabla \times \vec{H} = \frac{n^2}{c^2} \frac{\partial}{\partial t} (\nabla \times \vec{E}) \qquad \text{A-(14)}$$

or, by Fq. (7),

$$\nabla(\nabla \cdot \vec{H}) - (\nabla \cdot \nabla)\vec{H} = -\frac{n^2}{c^2} \frac{\partial^2 \vec{H}}{\partial t^2} \qquad \text{A-(15)}$$

Thus

$$\nabla^2 \vec{H} = \frac{n^2}{c^2} \frac{\partial^2 \vec{H}}{\partial t^2} \qquad \text{A-(16)}$$

A plane-wave solution to Fq. (13) is of the form

$$\vec{E} = \vec{E}_{oo} e^{i(\omega t - \vec{k} \cdot \vec{r})} \qquad \text{A-(17)}$$

$$= \vec{E}_{oo} e^{i(\omega t - k_x x - k_y y - k_z z)} \qquad \text{A-(18)}$$

$$\equiv \vec{E}_o(x, y) e^{i(\omega t - \beta z)} \qquad \text{A-(19)}$$

Here ω is the angular frequency and $\vec{r} = \hat{x}x + \hat{y}y + \hat{z}z$ ($\hat{x}$, $\hat{y}$, and $\hat{z}$ are unit normals). Also, $\vec{k}$ is the propagation vector which has a direction normal to the wavefront and a magnitude of nk_o, where n is the index of refraction and $k_o = \omega/c = 2\pi/\lambda_o$ (and λ_o is the free-space wavelength); β is the wavevector component along the $\hat{z}$-axis [as shown in Figure II-1] and has the value $k_o n_g \sin\theta$.

It may be noted that any x or y dependence has been absorbed into the vector $\vec{E}_o(x,y)$, where

$$\vec{E}_o(x,y) = \hat{x}\,E_{ox} + \hat{y}\,E_{oy} + \hat{z}\,E_{oz} \qquad \text{A-(20)}$$

A similar solution exists for H:

$$\vec{H} = \vec{H}_o(x,y)\,e^{i(\omega t - \beta z)} \qquad \text{A-(21)}$$

where

$$\vec{H}_o(x,y) = \hat{x}\,H_{ox} + \hat{y}\,H_{oy} + \hat{z}\,H_{oz} \qquad \text{A-(22)}$$

From the geometry of Figure II-1 for a planar (slab) waveguide, $\vec{E}_o$ and $\vec{H}_o$ have no y-dependence. Hence,

$$\vec{E}_o = \vec{E}_o(x) \;;\; \vec{E} = \vec{E}(x,z) \qquad \text{A-(23)}$$

and

$$\vec{H}_o = \vec{H}_o(x) \;;\; \vec{H} = \vec{H}(x,z) \qquad \text{A-(24)}$$

and all $\partial/\partial y$ terms in Eqs. (13) and (16) are zero.

Vector Potentials

Returning now to Eqs. (5) - (8), if the time dependence $e^{i\omega t}$ in Eqs. (19) and (21) is used, then the following set of equations

result (assuming n is not a function of time):

$$\nabla \cdot (n^2 \vec{E}) = 0 \qquad \text{A-(25)}$$

$$\nabla \cdot (\vec{H}) = 0 \qquad \text{A-(26)}$$

$$\nabla \times \vec{E} = -i\omega \vec{H}/c = -i k_0 \vec{H} \qquad \text{A-(27)}$$

$$\nabla \times \vec{H} = i\omega n^2 \vec{E}/c = i k_0 n^2 \vec{E} \qquad \text{A-(28)}$$

In view of the divergenceless character of $n^2\vec{E}$ and of $\vec{H}$, it is possible to express the fields in terms of a magnetic vector potential $\vec{A}$ or in terms of an electric vector potential $\vec{F}$.[1,2] Any vector field $\vec{H}$ whose divergence vanishes, as in Eq. (26), can be written as

$$\vec{H} = \nabla \times \vec{A} \qquad \text{A-(29)}$$

Substituting Eq. (29) into Eq. (27) produces

$$\nabla \times (\vec{E} + i k_0 \vec{A}) = 0 \qquad \text{A-(30)}$$

Any curl-free vector field can be expressed as the gradient of a scalar potential. Thus

$$\vec{E} + i k_0 \vec{A} = \nabla \psi^A \qquad \text{A-(31)}$$

where ψ^A is an electric scalar potential. Re-writing Eq. (31), $\vec{E}$ may be written as

$$\vec{E} = -ik_0\vec{A} + \nabla\psi^A \qquad \text{A-(32)}$$

In an exactly analogous fashion, Eq. (25) allows $n^2\vec{E}$ to be written as

$$n^2\vec{E} = \nabla \times (-n^2\vec{F}) \qquad \text{A-(33)}$$

Substituting Eq. (33) into Eq. (28) produces

$$\nabla \times (\vec{H} + ik_0 n^2 \vec{F}) = 0 \qquad \text{A-(34)}$$

which will be satisfied if

$$\vec{H} = -ik_0 n^2 \vec{F} + \nabla\psi^F \qquad \text{A-(35)}$$

Here, ψ^F is a magnetic scalar potential. Eqs. (32) and (35) are dual equations.

To obtain an equation for $\vec{A}$, Eqs. (29) and (32) are substituted into Eq. (28):

$$\nabla \times \nabla \times \vec{A} = k_0^2 n^2 \vec{A} + ik_0 n^2 \nabla\psi^A \qquad \text{A-(36)}$$

To obtain an equation for $\vec{F}$, Eq. (33) and (35) are substituted into Eq. (27):

$$\nabla \times \left[\frac{1}{n^2} \nabla \times (-n^2 \vec{F}) \right] = -k_o^2 n^2 \vec{F} - i k_o \nabla \psi^F \qquad \text{A-(37)}$$

Returning now to Eqs. (27), (28), (29), and (33), the following may be said. Solutions for $\vec{E}$ are given by Eq. (28) and also by Eq. (33); hence the superposition of these two partial solutions is also a solution, i.e.,

$$\vec{E} = -\frac{1}{n^2} \nabla \times (n^2 \vec{F}) - (i/k_o n^2) \nabla \times \vec{H} \qquad \text{A-(38)}$$

which, by Eq. (29), becomes

$$\vec{E} = -\frac{1}{n^2} \nabla \times (n^2 \vec{F}) - (i/k_o n^2) \nabla \times \nabla \times \vec{A} \qquad \text{A-(39)}$$

Similarly, a complete solution for $\vec{H}$ is given by the superposition of Eqs. (27) and (29):

$$\vec{H} = \nabla \times \vec{A} + (i/k_o) \nabla \times \vec{E} \qquad \text{A-(40)}$$

which, by Eq. (33), becomes

$$\vec{H} = \nabla \times \vec{A} - (i/k_o) \nabla \times \left(\frac{1}{n^2} \nabla \times n^2 \vec{F} \right) \qquad \text{A-(41)}$$

Homogeneous Waveguides

For homogeneous waveguides, the refractive index is not a function of position; hence Eq. (37) reduces to

$$-\nabla \times \nabla \times \vec{F} = -k_o^2 n^2 \vec{F} - i k_o \nabla \psi^F \qquad \text{A-(42)}$$

Using the vector identity

$$\nabla \times \nabla \times \vec{U} = \nabla(\nabla \cdot \vec{U}) - \nabla^2 \vec{U} \qquad \text{A-(43)}$$

Eqs. (36) and (42) can be written as

$$\nabla\left(\nabla \cdot \vec{A}\right) - \nabla^2 \vec{A} = k_o^2 n^2 \vec{A} + i k_o n^2 \nabla \psi^A \qquad \text{A-(44)}$$

$$-\nabla\left(\nabla \cdot \vec{F}\right) + \nabla^2 \vec{F} = -k_o^2 n^2 \vec{F} - i k_o \nabla \psi^F \qquad \text{A-(45)}$$

There is a good deal of arbitrariness in the choice of the vector potentials $\vec{A}$ and $\vec{F}$, and it is possible to specify the divergence of them. Thus letting

$$\nabla \cdot \vec{A} = i k_o n^2 \psi^A \qquad \text{A-(46)}$$

$$\nabla \cdot \vec{F} = i k_o \psi^F \qquad \text{A-(47)}$$

then Eqs. (44) and (45) reduce to the wave equations

$$\nabla^2 \vec{A} + k_o^2 n^2 \vec{A} = 0 \qquad \text{A-(48)}$$

$$\nabla^2 \vec{F} + k_o^2 n^2 \vec{F} = 0 \qquad \text{A-(49)}$$

For homogeneous waveguides, Eqs. (39) and (41) reduce to

$$\vec{E} = -\nabla \times \vec{F} - (i/k_o n^2) \nabla \times \nabla \times \vec{A} \qquad \text{A-(50)}$$

$$\vec{H} = \nabla \times \vec{A} - (i/k_o) \nabla \times \nabla \times \vec{F} \qquad \text{A-(51)}$$

which, by Eq. (43), become

$$\vec{E} = -\nabla \times \vec{F} - (i/k_o n^2)\left[\nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}\right] \qquad \text{A-(52)}$$

$$\vec{H} = \nabla \times \vec{A} - (i/k_o)\left[\nabla(\nabla \cdot \vec{F}) - \nabla^2 \vec{F}\right] \qquad \text{A-(53)}$$

At this point it is convenient to consider some particular choice of potentials. Letting

$$\vec{A} = 0 \qquad \text{A-(54)}$$

$$\vec{F} = \hat{x} F_x \qquad \text{A-(55)}$$

then the wave equations [Eqs. (48) and (49)] reduce to

$$\frac{\partial^2 F_x}{\partial z^2} + \frac{\partial^2 F_x}{\partial x^2} + k_o^2 n^2 F_x = 0 \qquad \text{A-(56)}$$

where the $\partial/\partial y$ term has been set equal to zero for the slab waveguide case. Further the equations for $\vec{E}$ and $\vec{H}$ [Eqs. (52) and (53)] become:

$$\vec{E} = -\nabla \times \vec{F} = -\begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & 0 & \frac{\partial}{\partial z} \\ F_x & 0 & 0 \end{vmatrix} \qquad \text{A-(57)}$$

$$\vec{H} = -(i/k_o)\left[\nabla\left(\frac{\partial F_x}{\partial x}\right) - \nabla^2 \vec{F}\right] \qquad \text{A-(58)}$$

Expanding Eqs. (57) and (58) into rectangular coordinates results in the following:

$$E_x = 0 \qquad \text{A-(59)}$$

$$E_y = -\frac{\partial F_x}{\partial z} \qquad \text{A-(60)}$$

$$E_z = 0 \qquad \text{A-(61)}$$

$$H_x = (i/k_o)\frac{\partial^2 F_x}{\partial z^2} \qquad \text{A-(62)}$$

$$H_y = 0 \qquad \text{A-(63)}$$

$$H_z = -(i/k_o)\frac{\partial^2 F_x}{\partial z \partial x} \qquad \text{A-(64)}$$

For this choice of $\vec{A}$ and $\vec{F}$ it may be seen that there is no $\hat{z}$-component of $\vec{E}$; for this reason this is called the transverse electric (TE) mode of propagation. Letting

$$\vec{A} = \hat{x} A_x \qquad \text{A-(65)}$$

$$\vec{F} = 0 \qquad \text{A-(66)}$$

then the wave equations reduce to

$$\frac{\partial^2 A_x}{\partial z^2} + \frac{\partial^2 A_x}{\partial x^2} + k_o^2 n^2 A_x = 0 \qquad \text{A-(67)}$$

From Eqs. (52) and (53), the following equations for $\vec{E}$ and $\vec{H}$ result

$$\vec{E} = -(i/k_0 n^2)\left[\nabla\left(\frac{\partial A_x}{\partial x}\right) - \nabla^2\vec{A}\right] \qquad \text{A-(68)}$$

$$\vec{H} = \nabla \times \vec{A} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial}{\partial x} & 0 & \frac{\partial}{\partial z} \\ A_x & 0 & 0 \end{vmatrix} \qquad \text{A-(69)}$$

or, in rectangular coordinates,

$$E_x = (i/k_0 n^2)\frac{\partial^2 A_x}{\partial z^2} \qquad \text{A-(70)}$$

$$E_y = 0 \qquad \text{A-(71)}$$

$$E_z = -(i/k_0 n^2)\frac{\partial^2 A_x}{\partial z \partial x} \qquad \text{A-(72)}$$

$$H_x = 0 \qquad \text{A-(73)}$$

$$H_y = \frac{\partial A_x}{\partial z} \qquad \text{A-(74)}$$

$$H_z = 0 \qquad \text{A-(75)}$$

For this choice of $\vec{A}$ and $\vec{F}$ there is no $\hat{z}$-component of $\vec{H}$; hence this is called the transverse magnetic (TM) mode of propagation.

Finally, for a situation where neither a pure TE nor a pure TM exists, the following choice of $\vec{A}$ and $\vec{F}$ is useful:

$$\vec{A} = \hat{x}\, A_x \qquad \text{A-(76)}$$

$$\vec{F} = \hat{x}\, F_x \qquad \text{A-(77)}$$

Then, from Eqs. (52) and (53)

$$E_x = \left(i/k_0 n^2\right) \frac{\partial^2 A_x}{\partial z^2} \qquad \text{A-(78)}$$

$$E_y = -\frac{\partial F_x}{\partial z} \qquad \text{A-(79)}$$

$$E_z = -\left(i/k_0 n^2\right) \frac{\partial^2 A_x}{\partial z \partial x} \qquad \text{A-(80)}$$

$$H_x = \left(i/k_0\right) \frac{\partial^2 F_x}{\partial z^2} \qquad \text{A-(81)}$$

$$H_y = \frac{\partial A_x}{\partial z} \qquad \text{A-(82)}$$

$$H_z = -\left(i/k_0\right) \frac{\partial^2 F_x}{\partial z\, \partial x} \qquad \text{A-(83)}$$

It may be seen that Eqs. (76) - (83) are nothing more than the superposition of Eqs. (60) - (65) [the TE case] and Eqs. (70) - (75) [the TM case]. Thus a field of any polarization in a homogeneous planar guide can be expressed as the sum of a TE field and a TM field.

Inhomogeneous Waveguides

Returning now to Eq. (37), use of the vector identity

$$\nabla \times (u\vec{U}) = \nabla u \times \vec{U} + u \nabla \times \vec{U} \qquad \text{A-(84)}$$

allows Eq. (37) to be written as

$$-\nabla \times \left[\frac{1}{n^2} \left(\nabla n^2 \times \vec{F} + n^2 \nabla \times \vec{F} \right) \right] = -k_0^2 n^2 \vec{F} - i k_0 \nabla \psi^F \qquad \text{A-(85)}$$

The assumption is now made that if a refractive index gradient exists, then it lies in the $\hat{x}$ direction only, i.e.,

$$n = n(x) \qquad \text{A-(86)}$$

Further, $\vec{F}$ and $\vec{A}$ are still chosen to lie in the $\hat{x}$ direction [i.e., $\vec{F} = \hat{x}F_x$, $\vec{A} = \hat{x}F_x$] as in the homogeneous case. As a consequence, the term $\nabla n^2 \times \vec{F}$ in Eq. (85) has the value

$$\begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \frac{\partial (n^2)}{\partial x} & 0 & 0 \\ F_x & 0 & 0 \end{vmatrix} = 0 \qquad \text{A-(87)}$$

Hence, Eq. (85) reduces to

$$-\nabla\times\nabla\times\vec{F} = -k_o^2 n^2 \vec{F} - ik_o \nabla\psi^F \qquad \text{A-(88)}$$

which is identical to Eq. (42) for the homogeneous case. Eq. (36) for $\vec{A}$ is still valid for the inhomogeneous case.

Using Eq. (43), Eqs. (36) and (88) can be written in the form of Eqs. (44) and (45). Then Eqs. (46) and (47) are again used as the choice for the divergence of the vector potentials $\vec{A}$ and $\vec{F}$. Substitution of Eqs. (46) and (47) into Eqs. (44) and (45) yield the following equations.

$$\nabla^2\vec{A} - ik_o \nabla\psi^A \nabla n^2 + k_o^2 n^2 \vec{A} = 0 \qquad \text{A-(89)}$$

$$\nabla^2\vec{F} + k_o^2 n^2 \vec{F} = 0 \qquad \text{A-(90)}$$

Since $\psi^A = (-i/k_o n^2)\nabla\cdot\vec{A}$, Eq. (89) becomes

$$\nabla^2\vec{A} - (1/n^2)(\nabla\cdot\vec{A})(\nabla n^2) + k_o^2 n^2 \vec{A} = 0 \qquad \text{A-(91)}$$

It may be noted that Eq. (91) for $\vec{A}$ for the inhomogeneous case is identical to Eq. (48) for the homogeneous case, except for the term $-(\nabla\cdot\vec{A})(\nabla n^2)$ which is zero for n = constant. Eq. (90) for $\vec{F}$ is identical to Eq. (49), i.e., $\vec{F}$ obeys the same wave equation for both the homogeneous and inhomogeneous cases. Using Eqs. (84) and (85), then Eqs. (90) and (91) become the wave equations

$$\frac{\partial^2 F_x}{\partial z^2} + \frac{\partial^2 F_x}{\partial x^2} + k_o^2 n^2(x) F_x = 0 \qquad \text{A-(92)}$$

$$\frac{\partial^2 A_x}{\partial z^2} + \frac{\partial^2 A_x}{\partial x^2} - \frac{1}{n^2(x)} \frac{\partial A_x}{\partial x} \frac{\partial [n^2(x)]}{\partial x} + k_o^2 n^2(x) A_x = 0 \qquad \text{A-(93)}$$

The equations for $\vec{E}$ and $\vec{H}$ are given by Eqs. (39) and (41). Employing the vector identies given in Eqs. (43) and (84), Eqs. (39) and (41) can be written as

$$\vec{E} = -\frac{1}{n^2}\left(\nabla n^2 + \vec{F} + n^2 \nabla \times \vec{F}\right) - \left(i/k_o n^2\right)\left[\nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}\right] \qquad \text{A-(94)}$$

$$\vec{H} = \nabla \times \vec{A} - \left(i/k_o\right)\nabla \times \left[\frac{1}{n^2}\left(\nabla n^2 \times \vec{F} + n^2 \nabla \times \vec{F}\right)\right] \qquad \text{A-(95)}$$

Since $\nabla n^2 \times \vec{F}$ is zero, Eqs. (86) and (87) become

$$\vec{E} = -\nabla \times \vec{F} - \left(i/k_o n^2\right)\left[\nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}\right] \qquad \text{A-(96)}$$

$$\vec{H} = \nabla \times \vec{A} - \left(i/k_o\right)\nabla \times \nabla \times \vec{F} \qquad \text{A-(97)}$$

$$= \nabla \times \vec{A} - \left(i/k_o\right)\left[\nabla(\nabla \cdot \vec{F}) - \nabla^2 \vec{F}\right] \qquad \text{A-(98)}$$

Again, using the choices $\vec{A} = \hat{x}A_x$ and $\vec{F} = \hat{x}F_x$, $\vec{E}$ and $\vec{H}$ in Eqs. (96) and (98) then become, in rectangular coordinates,

$$E_x = \left[i/k_0^2 n^2(x)\right] \frac{\partial^2 A_x}{\partial z^2} \qquad \text{A-(99)}$$

$$E_y = -\frac{\partial F_x}{\partial z} \qquad \text{A-(100)}$$

$$E_z = -\left[i/k_0 n^2(x)\right] \frac{\partial^2 A_x}{\partial z \partial x} \qquad \text{A-(101)}$$

$$H_x = \left(i/k_0\right) \frac{\partial^2 F_x}{\partial z^2} \qquad \text{A-(102)}$$

$$H_y = \frac{\partial A_x}{\partial z} \qquad \text{A-(103)}$$

$$H_z = -\left(i/k_0\right) \frac{\partial^2 F_x}{\partial z \partial x} \qquad \text{A-(104)}$$

It may be noted that Eqs. (99) and (104) are identical to Eqs. (86) - (91) if the refractive index n is a constant. Further, it was the choice $\vec{A} = \hat{x}A_x$ and $\vec{F} = \hat{x}F_x$ (i.e., that $\vec{A}$ and $\vec{F}$ lie in the direction of the index gradient) that resulted in considerable simplification of the inhomogeneous waveguide equations, especially Eqs. (85), (94) and (95).

In the homogeneous case, the choice $\vec{A} = 0$, $\vec{F} = \hat{x}F_x$ led to Eqs. (60) - (65), the TE mode of propagation. So too for the inhomogeneous case. If $\vec{A} = 0$, then E_x, E_z, and H_y in Eqs. (93) - (98) are identically zero, producing a pure TE mode. If $\vec{F}$ is set equal to zero, then E_y, H_x, and H_z in Eqs. (101) - (104) are identically zero, producing a pure TM mode. Further, as with the homogeneous case, for the situation where neither a pure TE nor a pure TM exists, by choosing $\vec{A}$ and $\vec{F}$ to be both non-zero, a field of any polarization in an inhomogeneous planar guide can be expressed as the sum of a TE field and a TM field.

APPENDIX B

HOMOGENEOUS ASYMMETRIC WAVEGUIDES

TE Modes

In order to obtain solutions for $\vec{E}$ and $\vec{H}$ for the TE case, the following choice is made for $F_x[\vec{F} = \hat{x}F_x]$:

$$F_{xa} = \left(F_1/i\beta_a\right) e^{i(\omega t - \beta_a z)} e^{-ax} \qquad \text{for } x \geq 0 \qquad \text{B-(1)}$$

$$F_{xg} = \left(1/i\beta_g\right) e^{i(\omega t - \beta_g z)} \left[F_2 e^{igx} + F_3 e^{-igx}\right] \qquad \text{for } -W \leq x \leq 0 \qquad \text{B-(2)}$$

$$F_{xs} = \left(F_4/i\beta_s\right) e^{i(\omega t - \beta_s z)} e^{sx} \qquad \text{for } x \leq -W \qquad \text{B-(3)}$$

where F_1, F_2, F_3, and F_4 are constants (possibly complex). The choice for $\vec{F}$ given in Eqs. (1) - (3), when substituted into Eqs. A-(59) through A-(64), insures that the wave is confined to the n_g region (the dielectric waveguide) and that the field goes to zero as $x \rightarrow \pm \infty$. It is also consistent with the form of $\vec{E}$ and $\vec{H}$ given by Eqs. A-(19) and A-(21).

Substituting Eqs. (1) - (3) into Eq. A-(56), the wave equation for $\vec{F}$, yields

$$\left(a^2-\beta_a^2+k_o^2 n_a^2\right)F_{xa}=0 \qquad \text{B-(4)}$$

$$\left(1/i\beta_g\right)e^{i(\omega t-\beta_3 z)}\left(-g^2-\beta_g^2+k_o^2 n_g^2\right)\left(F_2 e^{igx}+F_3 e^{-igx}\right)=0 \qquad \text{B-(5)}$$

$$\left(s^2-\beta_s^2+k_o^2 n_s^2\right)F_{xs}=0 \qquad \text{B-(6)}$$

Since e^{igx} and e^{-igx} are independent functions, Eq. (5) can be re-written as two equations:

$$\left(F_2 e^{igx}\right)\left(-g^2-\beta_g^2+k_o^2 n_g^2\right)=0 \qquad \text{B-(7)}$$

$$\left(F_3 e^{-igx}\right)\left(-g^2-\beta_g^2+k_o^2 n_g^2\right)=0 \qquad \text{B-(8)}$$

Re-expressing Eqs. (4) - (8) leads to the following:

$$\beta_a^2=k_o^2 n_a^2+a^2 \qquad \text{B-(9)}$$

$$\beta_g^2 = k_o^2 n_g^2 - g^2 \qquad \text{B-(10)}$$

$$\beta_s^2 = k_o^2 n_s^2 + s^2 \qquad \text{B-(11)}$$

It is possible to determine the relationships among β_a, β_g, and β_s by imposing the standard boundary conditions at the dielectric interfaces, namely that the components of $\vec{E}$ and $\vec{H}$ (F_y and H_z) which are tangential to the $n_a - n_g$ and $n_g - n_s$ boundaries be continuous across the boundaries.

$$\left.\begin{aligned} [E_y]_a &= [E_y]_g \qquad \text{B-(12)} \\ [H_z]_a &= [H_z]_g \qquad \text{B-(13)} \end{aligned}\right\} \text{at } x=0$$

$$\left.\begin{aligned} [E_y]_s &= [E_y]_g \qquad \text{B-(14)} \\ [H_z]_s &= [H_z]_g \qquad \text{B-(15)} \end{aligned}\right\} \text{at } x=-W$$

However, Eqs. A-(60) and A-(64) imply that the boundary conditions on H_z are equivalent to the requirement that $(i/k_o)\partial E_y/\partial x$ be continuous across the boundaries. Thus Eqs. (13) and (15) can be re-written as

$$\left[\frac{\partial E_y}{\partial x}\right]_a = \left[\frac{\partial E_y}{\partial x}\right]_g \quad \text{at } x=0 \qquad \text{B-(16)}$$

$$\left[\frac{\partial E_y}{\partial x}\right]_s = \left[\frac{\partial E_y}{\partial x}\right]_g \quad \text{at } x=-W \qquad \text{B-(17)}$$

From Eqs. (1), (2), and A-(60), Eq. (12) implies that, at the n_a - n_g boundary,

$$i\beta_a F_{xa}\Big|_{x=0} = i\beta_g F_{xg}\Big|_{x=0} \qquad \text{B-(18)}$$

or,

$$F_1 e^{i(\omega t-\beta_a z)} = (F_2+F_3) e^{i(\omega t-\beta_g z)} \qquad \text{B-(19)}$$

hence,

$$F_1 e^{-i\beta_a z} = (F_2+F_3) e^{-i\beta_g z} \qquad \text{B-(20)}$$

Since F_1, F_2, and F_3 are constants, for arbitrary z the only possible relationship between β_a and β_g is

$$\beta_a = \beta_g \qquad \text{B-(21)}$$

Similarly, at the n_g - n_s boundary, Eqs. (2), (3), (14) and A-(60) imply that

$$i\beta_s F_{xs}\Big|_{x=-W} = i\beta_g F_{xg}\Big|_{x=-W} \qquad \text{B-(22)}$$

or,

$$F_4 e^{-i\beta_s z} e^{-sW} = e^{-i\beta_g z}\left[F_2 e^{-igW} + F_3 e^{igW}\right] \qquad \text{B-(23)}$$

Since $F_4 e^{-sW}$, $F_2 e^{-igW}$, and $F_3 e^{igW}$ are constants, the following relationship must exist:

$$\beta_g = \beta_s \qquad \text{B-(24)}$$

Recalling that β is the wavevector component along the $\hat{z}$-axis and that, for the n_g region (the waveguide), $\beta = k_o n_g \sin\theta$, then it is convenient to let

$$\beta \equiv k_o n_g \sin\theta = \beta_a = \beta_g = \beta_s \qquad \text{B-(25)}$$

where θ is defined in Figure II-1. Eq. (25) thus implies that the wave propagates in the $\hat{z}$-direction at the same velocity v_z in the n_a, n_g, and n_s regions, since $v_z = c\beta$.

Eqs. (9) - (11) can now be transformed with the aid of Eq. (25):

$$a = k_o \left(n_g^2 \sin^2\theta - n_a^2 \right)^{1/2} \qquad \text{B-(26)}$$

$$g = k_o \left(n_g^2 - n_g^2 \sin^2\theta \right)^{1/2} \qquad \text{B-(27)}$$

$$= k_o n_g \cos\theta \qquad \text{B-(28)}$$

$$s = k_o \left(n_g^2 \sin^2\theta - n_s^2 \right)^{1/2} \qquad \text{B-(29)}$$

Use of Eqs. A-(60), and A-(64) allows Eqs. (12), (16), (14), and (17) to be written respectively as

$$\left[F_{xa} \right]_{x=0} = \left[F_{xg} \right]_{x=0} \qquad \text{B-(30)}$$

$$\left[\frac{\partial F_{xa}}{\partial x}\right]_{x=0} = \left[\frac{\partial F_{xg}}{\partial x}\right]_{x=0} \qquad \text{B-(31)}$$

$$\left[F_{xs}\right]_{x=-W} = \left[F_{xg}\right]_{x=-W} \qquad \text{B-(32)}$$

$$\left[\frac{\partial F_{xs}}{\partial x}\right]_{x=-W} = \left[\frac{\partial F_{xg}}{\partial x}\right]_{x=-W} \qquad \text{B-(33)}$$

From Eqs. (1) - (3) and (25), Eqs. (30) - (33) become

$$F_1 = F_2 + F_3 \qquad \text{B-(34)}$$

$$-a F_1 = ig F_2 - ig F_3 \qquad \text{B-(35)}$$

$$F_4 e^{-sW} = F_2 e^{-igW} + F_3 e^{igW} \qquad \text{B-(36)}$$

$$s F_4 e^{-sW} = ig F_2 e^{-igW} - ig F_3 e^{igW} \qquad \text{B-(37)}$$

Multiplying Eq. (34) by +a and adding it to Eq. (35) produces

$$F_2 (a+ig) = -F_3 (a-ig) \qquad \text{B-(38)}$$

Multiplying Eq. (36) by -s and adding it to Eq. (37) produces

$$F_2(-s+ig)e^{-igW} = F_3(s+ig)e^{igW} \qquad \text{B-(39)}$$

Dividing Eq. (38) by Eq. (39) results in

$$\frac{a+ig}{s-ig}e^{igW} = \frac{a-ig}{s+ig}e^{-igW} \qquad \text{B-(40)}$$

or

$$e^{2igW} = \frac{(a-ig)}{(a+ig)}\frac{(s-ig)}{(s+ig)} \qquad \text{B-(41)}$$

From complex algebra

$$Z = X + iY = +i(-iX+Y) = +i\sqrt{X^2+Y^2}\,e^{-i\tan^{-1}(X/Y)} \qquad \text{B-(42)}$$

$$Z^* = X - iY = -i(+iX+Y) = -i\sqrt{X^2+Y^2}\,e^{+i\tan^{-1}(X/Y)} \qquad \text{B-(43)}$$

and

$$\frac{Z^*}{Z} = \frac{X-iY}{X+iY} = -e^{2i\tan^{-1}(X/Y)} \qquad \text{B-(44)}$$

Thus, Eq. (41) can be re-written as

$$e^{2igW} = e^{2i\tan^{-1}(a/g)}\ e^{2i\tan^{-1}(s/g)} \qquad \text{B-(45)}$$

With no loss of generality, e^{2igW} can be written as $e^{2igW}e^{-(2im\pi)}$ where m = 0, 1, 2, ... Thus

$$\exp\left[i\,(2gW - 2m\pi)\right] = \exp\left\{i\left[2\tan^{-1}(a/g) + 2\tan^{-1}(s/g)\right]\right\} \qquad \text{B-(46)}$$

or

$$2gW - 2\varphi_{ag}^{TE} - 2\varphi_{gs}^{TE} = 2m\pi \qquad \text{B-(47)}$$

where

$$\varphi_{ag}^{TE} \equiv \tan^{-1}(a/g) \qquad \text{B-(48)}$$

$$\varphi_{gs}^{TE} \equiv \tan^{-1}(s/g) \qquad \text{B-(49)}$$

Thus, from Eqs. (26), (28), and (29),

$$\varphi_{ag}^{TE} = \tan^{-1}\left[\left(n_g^2 \sin^2\theta - n_a^2\right)^{1/2} / n_g \cos\theta\right] \qquad \text{B-(50)}$$

$$\varphi_{gs}^{TE} = \tan^{-1}\left[\left(n_g^2 \sin^2\theta - n_s^2\right)^{1/2} / n_g \cos\theta\right] \qquad \text{B-(51)}$$

Using Eq. (28), Eq. (47) can be written as

$$2k_o W n_g \cos\theta - 2\varphi_{ag}^{TE} - 2\varphi_{gs}^{TE} = 2m\pi \qquad \text{B-(52)}$$

It may be noted that Eq. (52), the waveguide equation, is a transcendental equation in θ.

Eq. (52) is useful for calculating the minimum thickness $(W_{min})_{TE}$ necessary for propagating a given mode of order m. Solving for W gives:

$$W = \left(\varphi_{ag}^{TE} + \varphi_{gs}^{TE} + m\pi\right) / k_o n_g \cos\theta \qquad \text{B-(53)}$$

When $\sin\theta = n_s/n_g \equiv \sin\theta_{gs}$, then θ is at the critical angle θ_{gs} for total internal reflection at the $n_g - n_s$ boundary. Since $n_s > n_a$, total reflection still occurs at the $n_a - n_g$ boundary. However, if $\theta < \theta_{gs}$, plane waves escape or leak out into the n_s medium and propagation ceases, i.e., the mode is cut off. The cutoff angle θ_{cc} is therefore

$$\sin\theta_{cc} \equiv \sin\theta_{gs} = n_s / n_g \qquad \text{B-(54)}$$

By Eqs. (50) and (54), at cutoff,

$$\varphi_{ag}^{TE} = \tan^{-1}\left[\left(n_s^2 - n_a^2\right)^{1/2} / n_g \left(1 - n_s^2/n_g^2\right)^{1/2}\right] \qquad \text{B-(55)}$$

since

$$\cos\theta_{gs} = \left(1 - \sin^2\theta_{gs}\right)^{1/2} = \left(1 - n_s^2/n_g^2\right)^{1/2} \qquad \text{B-(56)}$$

Thus, for TE modes at cutoff,

$$\varphi_{ag}^{TE} = \tan^{-1}\left[\left(n_s^2 - n_a^2\right)^{1/2} / \left(n_g^2 - n_a^2\right)^{1/2}\right] \qquad \text{B-(57)}$$

Further, by Eqs. (51) and (54), for TE modes at cutoff,

$$\varphi_{gs}^{TE} = 0 \qquad \text{B-(58)}$$

Since $1/k_o n_g \cos\theta_{gs} = \lambda_o/2\pi(n_g^2 - n_s^2)^{1/2}$, Eq. (53) may be written for TE modes as

$$(W_{min})_{TE} = \frac{\lambda_o}{2\pi(n_g^2-n_s^2)^{1/2}}\left\{m\pi + \tan^{-1}\left[\frac{(n_s^2-n_a^2)^{1/2}}{(n_g^2-n_s^2)^{1/2}}\right]\right\} \quad \text{B-(59)}$$

In order to find the final solutions for $\vec{E}$ and $\vec{H}$, it is necessary to calculate the coefficients F_1, F_2, F_3, and F_4, one of which is arbitrary, say F_2. Using Eq. (38), it may be seen that F_2 and F_3 differ only by a phase factor. Thus letting

$$F_3 = e^{i2\phi^{TE}} F_2 \quad \text{B-(60)}$$

then Eqs. (38) and (39) become

$$(a+ig) = -e^{i2\phi^{TE}}(a-ig) \quad \text{B-(61)}$$

$$(s-ig)\,e^{-igW} = -e^{i2\phi^{TE}}(s+ig)e^{igW} \quad \text{B-(62)}$$

Taking the complex conjugate of Eq. (62) and dividing it into Eq. (61) produces

$$e^{-2igW}\,\frac{(a+ig)}{(s+ig)} = e^{i4\phi^{TE}}\,\frac{(a-ig)}{(s-ig)} \qquad \text{B-(63)}$$

or,

$$e^{i4\phi^{TE}} = \frac{(a+ig)}{(a-ig)}\,\frac{(s-ig)}{(s+ig)}\,e^{-2igW} \qquad \text{B-(64)}$$

Using Eqs. (44), (48), and (49), Eq. (64) may be written as

$$e^{i4\phi^{TE}} = e^{-2i\phi_{ag}^{TE}}\; e^{+2i\phi_{gs}^{TE}}\; e^{-2igW}\; e^{2im\pi} \qquad \text{B-(65)}$$

Using Eq. (47), the exponentials in Eq. (65) yield the following relationship:

$$\phi^{TE} = -\,\phi_{ag}^{TE} \qquad \text{B-(66)}$$

Turning now to F_1 and F_4, Eqs. (34) and (36) can be solved in terms of F_2. Using Eqs. (60) and (66), Eq. (34) becomes

$$F_1 = F_2\left(1+e^{-2\varphi_{ag}^{TE}}\right) \qquad \text{B-(67)}$$

$$=2F_2\, e^{-i\varphi_{ag}^{TE}} \cos\left(\varphi_{ag}^{TE}\right) \qquad \text{B-(68)}$$

From Eq. (36)

$$F_4 = e^{sW}\left[F_2 e^{-igW} + F_3 e^{igW}\right] \qquad \text{B-(69)}$$

$$=e^{sW} F_2\, e^{-i\varphi_{ag}^{TE}}\left[e^{-i(gW-\varphi_{ag}^{TE})} + e^{i(gW-\varphi_{ag}^{TE})}\right] \qquad \text{B-(70)}$$

or, from Eq. (47),

$$F_4 = e^{sW} F_2 e^{-i\varphi_{ag}^{TE}}\left[e^{-i(\varphi_{gs}^{TE}+m\pi)} + e^{i(\varphi_{gs}^{TE}+m\pi)}\right] \qquad \text{B-(71)}$$

$$= 2F_2\, e^{sW}\, e^{-i\varphi_{ag}^{TE}} \cos\left(\varphi_{gs}^{TE}+m\pi\right) \qquad \text{B-(72)}$$

Finally, from Eqs. (2), (60), and (66),

$$F_2 e^{igx} + F_3 e^{-igx} = F_2 e^{-i\varphi_{ag}^{TE}} \left[e^{i(gx + \varphi_{ag}^{TE}} + e^{-i(gx + \varphi_{ag}^{TE})} \right] \qquad \text{B-(73)}$$

$$= 2F_2 e^{-i\varphi_{ag}^{TE}} \cos(gx + \varphi_{ag}^{TE}) \qquad \text{B-(74)}$$

It is now possible to write down the TE solutions for $\vec{E}$ and $\vec{H}$ by using Eqs. A-(59) through A-(64) together with Eqs. (1) - (3), (68), (72), and (74).

$$\vec{E}_{TE} = \hat{y} E_y \qquad \text{B-(75)}$$

$$\vec{H}_{TE} = \hat{x} H_x + \hat{z} H_z \qquad \text{B-(76)}$$

$$E_y(x,z,t) = E_{oy}(x) e^{i(\omega t - \beta z)} \qquad \text{B-(77)}$$

$$H_x(x,z,t) = H_{ox}(x) e^{i(\omega t - \beta z)} \qquad \text{B-(78)}$$

$$H_z(x,z,t) = H_{oz}(x) e^{i(\omega t - \beta z)} \qquad \text{B-(79)}$$

where

$$E_{oy}(x) = 2F_2 e^{-i\varphi_{ag}^{TE}} \cos(\varphi_{ag}^{TE}) e^{-ax} \quad \text{for } x \geqq 0 \qquad \text{B-(80)}$$

$$E_{oy}(x) = 2F_2 e^{-i\varphi_{ag}^{TE}} \cos(gx + \varphi_{ag}^{TE}) \quad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(81)}$$

$$E_{oy}(x) = 2F_2 e^{-i\varphi_{ag}^{TE}} \cos(\varphi_{gs}^{TE} + m\pi) e^{s(W+x)} \quad \text{for } x \leqq -W \qquad \text{B-(82)}$$

From Eqs. A-(60) and A-(62), $H_x = (-i/k_o)\partial E_y/\partial z$ for all x; hence from Eqs. (1) - (3)

$$H_{ox}(x) = -(\beta/k_o) E_{oy} \quad \text{for all } x \qquad \text{B-(83)}$$

Last of all, from Eqs. A-(60) and A-(64), $H_z = (i/k_o)\partial E_y/\partial x$; hence

$$H_{oz}(x) = -i(a/k_o) E_{oy}(x) \quad \text{for } x \geqq 0 \qquad \text{B-(84)}$$

$$H_{oz}(x) = -i(g/k_o)\tan(gx + \varphi_{ag}^{TE}) E_{oy} \quad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(85)}$$

$$H_{oz}(x) = +i(s/k_o)\,E_{oy}(x) \qquad \text{for } x \leqq -W \qquad \text{B-(86)}$$

TM Modes

Since the wave equation [Eq. A-(67)] for $\vec{A}$ has the same form as Eq. A-(56) for $\vec{F}$, the following choice is made for A_x [$\vec{A} = \hat{x}\,A_x$]:

$$A_{xa} = \left(A_1/i\beta_a\right) e^{i(\omega t - \beta_a z)}\, e^{-a_1 x} \qquad \text{for } x \geqq 0 \qquad \text{B-(87)}$$

$$A_{xg} = \left(1/i\beta_g\right) e^{i(\omega t - \beta_g z)}\left[A_2 e^{ig_1 x} + A_3 e^{-ig_1 x}\right] \qquad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(88)}$$

$$A_{xs} = \left(A_4/i\beta_s\right) e^{i(\omega t - \beta_s z)}\, e^{s_1 x} \qquad \text{for } x \leqq -W \qquad \text{B-(89)}$$

where A_1, A_2, A_3 and A_4 are constants (possibly complex). Substituting Eqs. (87) - (89) into Eq. A-(67) leads to the following equations:

$$\beta_a^2 = k_o^2 n_a^2 + a_1^2 \qquad \text{B-(90)}$$

$$\beta_g^2 = k_o^2 n_g^2 - g_1^2 \qquad \text{B-(91)}$$

$$\beta_s^2 = k_o^2 n_s^2 + s_1^2 \qquad \text{B-(92)}$$

Again it is possible to determine the relationships among β_a, β_g, and β_s by imposing the boundary conditions on $\vec{E}$ and $\vec{H}$. For TM modes, continuity of the tangential components of $\vec{E}$ and $\vec{H}$ (i.e., E_z and H_y) are given by

$$\left.\begin{aligned} [E_z]_a &= [E_z]_g \qquad \text{B-(93)} \\ [H_y]_a &= [H_y]_g \qquad \text{B-(94)} \end{aligned}\right\} \text{ at } x=0$$

$$\left.\begin{aligned} [E_z]_s &= [E_z]_g \qquad \text{B-(95)} \\ [H_y]_s &= [H_y]_g \qquad \text{B-(96)} \end{aligned}\right\} \text{ at } x=-W$$

However, Eqs. A-(72) and A-(74) imply that the boundary conditions on E_z are equivalent to the requirement that $(-i/k_o n^2)\partial H_y/\partial x$ be continuous across the boundaries. Thus, Eqs. (93) and (95) can be written as

$$\frac{1}{n_a^2}\left[\frac{\partial H_y}{\partial x}\right]_a = \frac{1}{n_g^2}\left[\frac{\partial H_y}{\partial x}\right]_g \quad \text{at } x=0 \qquad \text{B-(97)}$$

$$\frac{1}{n_s^2}\left[\frac{\partial H_y}{\partial x}\right]_s = \frac{1}{n_g^2}\left[\frac{\partial H_y}{\partial x}\right]_g \quad \text{at } x=-W \qquad \text{B-(98)}$$

Using Eqs. (87) and (88), Eq. (94) implies that at the n_a - n_g boundary

$$-i\beta_a A_{xa}\Big|_{x=0} = -i\beta_g A_{xg}\Big|_{x=0} \qquad \text{B-(99)}$$

or, from Eqs. (87) and (88),

$$A_1 e^{-i\beta_a z} = (A_2 + A_3)\, e^{-i\beta_g z} \qquad \text{B-(100)}$$

Since A_1, A_2, and A_3 are constants, for arbitrary z the following relationship must exist:

$$\beta_a = \beta_g \qquad \text{B-(101)}$$

Similarly, at the n_g - n_s boundary, the boundary conditions require that

$$-i\beta_s A_{xs}\Big|_{x=-W} = -i\beta_g A_{xg}\Big|_{x=-W} \qquad \text{B-(102)}$$

or

$$A_4 e^{-i\beta_s z} e^{-s_1 W} = e^{-i\beta_g z}\left[A_2 e^{-ig_1 W} + A_3 e^{ig_1 W}\right] \qquad \text{B-(103)}$$

Eq. (103) can be satisfied only if

$$\beta_g = \beta_s \qquad \text{B-(104)}$$

Again, since $\beta = k_o n_g \sin\theta$ is the wavevector component along the z-axis within the guide, it is convenient to let

$$\beta \equiv k_o n_g \sin\theta = \beta_a = \beta_g = \beta_s \qquad \text{B-(105)}$$

Substituting Eq. (105) into Eqs. (90) - (92), and solving for a_1, g_1, and s_1, leads to the following:

$$a_1 = k_o\left(n_g^2 \sin^2\theta - n_a^2\right)^{1/2} \qquad \text{B-(106)}$$

$$g_1 = k_o n_g \cos\theta \qquad \text{B-(107)}$$

$$s_1 = k_0 \left(n_g^2 \sin^2\theta - n_s^2\right)^{1/2} \qquad \text{B-(108)}$$

Comparing Eqs. (106) - (108) with Eqs. (26) - (29) for the TE case, it may be seen that

$$a_1 = a; \quad g_1 = g \; ; \quad s_1 = s \qquad \text{B-(109)}$$

Also, Eqs. (105), A-(72), and A-(74) allow Eqs. (97), (94), (98), and (96) to be written respectively as

$$\frac{1}{n_a^2}\left[\frac{\partial A_{xa}}{\partial x}\right]_{x=0} = \frac{1}{n_g^2}\left[\frac{\partial A_{xg}}{\partial x}\right]_{x=0} \qquad \text{B-(110)}$$

$$\left[A_{xa}\right]_{x=0} = \left[A_{xg}\right]_{x=0} \qquad \text{B-(111)}$$

$$\frac{1}{n_s^2}\left[\frac{\partial A_{xs}}{\partial x}\right]_{x=-W} = \frac{1}{n_g^2}\left[\frac{\partial A_{xg}}{\partial x}\right]_{x=-W} \qquad \text{B-(112)}$$

$$\left[A_{xs}\right]_{x=-W} = \left[A_{xs}\right]_{x=-W} \qquad \text{B-(113)}$$

Substituting Eqs. (87) - (89) and (109) into Eqs. (110) - (113) produces the following boundary relations:

$$-n_a^{-2} a A_1 = i n_g^{-2} g A_2 - i n_g^{-2} g A_3 \qquad \text{B-(114)}$$

$$A_1 = A_2 + A_3 \qquad \text{B-(115)}$$

$$-n_s^{-2} s A_4 e^{-sW} = i n_g^{-2} g A_2 e^{-igW} - i n_g^{-2} g A_3 e^{igW} \qquad \text{B-(116)}$$

$$A_4 e^{-sW} = A_2 e^{-igW} + A_3 e^{igW} \qquad \text{B-(117)}$$

Multiplying Eq. (115) by $+an_a^{-2}$ and adding it to Eq. (114) produces

$$-A_2 (a n_a^{-2} + i n_g^{-2} g) = A_3 (a n_a^{-2} - i n_g^{-2} g) \qquad \text{B-(118)}$$

Similarly, multiplying Eq. (117) by $-sn_s^{-2}$ and adding it to Eq. (116) produces

$$-A_2 (-s n_s^{-2} + i n_g^{-2} g) e^{-igW} = A_3 (-s n_s^{-2} - i n_g^{-2} g) e^{igW} \qquad \text{B-(119)}$$

Dividing Eq. (118) by Eq. (119) results in

$$\frac{(a n_a^{-2} + i n_g^{-2} g)}{(-s n_s^{-2} + i n_g^{-2} g)} e^{igW} = \frac{(a n_a^{-2} - i n_g^{-2} g)}{(-s n_s^{-2} - i n_g^{-2} g)} e^{-igW} \qquad \text{B-(120)}$$

or,

$$e^{2igW} = \frac{(a n_a^{-2} - ig n_g^{-2})}{(a n_a^{-2} + ig n_g^{-2})} \frac{(s n_s^{-2} - ig n_g^{-2})}{(s n_s^{-2} + ig n_g^{-2})} \qquad \text{B-(121)}$$

Using Eq. (44), Eq. (121) can be re-written as

$$e^{2igW} = e^{2i\tan^{-1}(a n_a^{-2}/g n_g^{-2})} e^{2i\tan^{-1}(s n_s^{-2}/g n_g^{-2})} \qquad \text{B-(122)}$$

Again, with no loss of generality, e^{2igW} can be written as $e^{2i(gW-m\pi)}$, where m = 0, 1, 2, ... Thus

$$\exp\left[i(2gW - 2m\pi)\right] = \exp\left[2i(\varphi_{ag}^{TM} + \varphi_{gs}^{TM})\right] \qquad \text{B-(123)}$$

or,

$$2gW - 2\phi_{ag}^{TM} - 2\phi_{gs}^{TM} = 2m\pi \qquad \text{B-(124)}$$

where

$$\phi_{ag}^{TM} \equiv \tan^{-1}\left[(a/g)\,(n_g^2/n_a^2)\right] \qquad \text{B-(125)}$$

$$\phi_{gs}^{TM} \equiv \tan^{-1}\left[(s/g)\,(n_g^2/n_s^2)\right] \qquad \text{B-(126)}$$

Then, from Eqs. (106) - (109),

$$\phi_{ag}^{TM} = \tan^{-1}\left\{(n_g^2/n_a^2)\left[(n_g^2\sin^2\theta - n_a^2)^{1/2}/n_g\cos\theta\right]\right\} \qquad \text{B-(127)}$$

$$\phi_{gs}^{TM} = \tan^{-1}\left\{(n_g^2/n_s^2)\left[(n_g^2\sin^2\theta - n_s^2)^{1/2}/n_g\cos\theta\right]\right\} \qquad \text{B-(128)}$$

It may be noted that ϕ_{ag}^{TM} and ϕ_{gs}^{TM} are identical to ϕ_{ag}^{TE} and ϕ_{gs}^{TE} in Eqs. (50) and (51) except for the factors n_g^2/n_a^2 and n_g^2/n_s^2 in the argument of the arctangent. Using Eqs. (107) and (109), Eq. (124) becomes

$$2k_0 W n_g\cos\theta - 2\phi_{ag}^{TM} - 2\phi_{gs}^{TM} = 2m\pi \qquad \text{B-(129)}$$

which is the waveguide equation for the TM case.

Solving for W and following the TE arguments (which apply equally well to TM modes) preceding Eq. (54) for the cutoff condition, the following equations result:

$$\phi_{ag}^{TM} = \tan^{-1}\left\{\left(n_g^2/n_a^2\right)\left[\left(n_s^2-n_a^2\right)/\left(n_g^2-n_s^2\right)\right]^{1/2}\right\} \qquad \text{B-(130)}$$

$$\phi_{gs}^{TM} = 0 \qquad \text{B-(131)}$$

so that the following equation is obtained for $(W_{min})_{TM}$:

$$(W_{min})_{TM} = \frac{\lambda_0}{2\pi\left(n_g^2-n_s^2\right)^{1/2}}\left\{m\pi + \tan^{-1}\left[\frac{n_g^2}{n_s^2}\frac{\left(n_s^2-n_a^2\right)^{1/2}}{\left(n_g^2-n_s^2\right)^{1/2}}\right]\right. \qquad \text{B-(132)}$$

Eq. (132) is similar to Eq. (59) except for the factor n_g^2/n_a^2.

In order to find the final solutions for $\vec{E}$ and $\vec{H}$ for the TM case, it is necessary to solve for A_1, A_3, and A_4 in terms of the arbitrary amplitude A_2, similar to what was done for the TE case. Letting

$$A_3 = e^{i2\varphi^{TM}} A_2 \qquad \text{B-(133)}$$

then Eqs. (118) and (119) become

$$-\left(a n_a^{-2} + i n_g^{-2} g\right) = e^{i2\varphi^{TM}} \left(a n_a^{-2} - i n_g^{-2} g\right) \qquad \text{B-(134)}$$

$$-\left(s n_s^{-2} - i n_g^{-2} g\right) e^{-igW} = e^{i2\varphi^{TM}} \left(s n_s^{-2} + i n_g^{-2} g\right) e^{igW} \qquad \text{B-(135)}$$

Taking the complex conjugate of Eq. (135) and dividing it into Eq. (134) produces

$$e^{-2igW} \frac{\left(a n_a^{-2} + i g n_g^{-2}\right)}{\left(s n_s^{-2} + i g n_g^{-2}\right)} = e^{i4\varphi^{TM}} \frac{\left(a n_a^{-2} - i g n_g^{-2}\right)}{\left(s n_s^{-2} - i g n_g^{-2}\right)} \qquad \text{B-(136)}$$

or

$$e^{i4\varphi^{TM}} = \frac{\left(a n_a^{-2} + i g n_g^{-2}\right)}{\left(a n_a^{-2} - i g n_g^{-2}\right)} \frac{\left(s n_s^{-2} - i g n_g^{-2}\right)}{\left(s n_s^{-2} + i g n_g^{-2}\right)} \qquad \text{B-(137)}$$

Using Eqs. (44), (125), and (126), Eq. (137) becomes

$$e^{i4\phi^{TM}} = e^{-2i\phi_{ag}^{TM}} e^{+2i\phi_{gs}^{TM}} e^{-2igW} \qquad \text{B-(138)}$$

or, with no loss of generality,

$$e^{i4\phi^{TM}} = e^{2i\left(-\phi_{ag}^{TM} + \phi_{gs}^{TM} + m\pi\right)} e^{-2igW} \qquad \text{B-(139)}$$

Using Eq. (124), the following relationship exists for the phase factor:

$$\phi^{TM} = -\phi_{ag}^{TM} \qquad \text{B-(140)}$$

Turning now to A_1 and A_4, Eqs. (115) and (117) can be solved in terms of A_2.

$$A_1 = A_2\left(1 + e^{-2\phi_{ag}^{TM}}\right) \qquad \text{B-(141)}$$

$$= 2A_2 e^{-i\phi_{ag}^{TM}} \cos\left(\phi_{ag}^{TM}\right) \qquad \text{B-(142)}$$

Also,

$$A_4 = e^{sW}\left[A_2 e^{-igW} + A_3 e^{igW}\right] \qquad \text{B-(143)}$$

$$= e^{sW} A_2 e^{-i\phi_{ag}^{TM}}\left[e^{-i(gW-\phi_{ag}^{TM})} + e^{i(gW-\phi_{ag}^{TM})}\right] \qquad \text{B-(144)}$$

or, from Eq. (124),

$$A_4 = e^{sW} A_2 e^{-i\phi_{ag}^{TM}}\left[e^{-i(\phi_{gs}^{TM}+m\pi)}\, e^{i(\phi_{gs}^{TM}+m\pi)}\right] \qquad \text{B-(145)}$$

$$= 2A_2 e^{sW} e^{-i\phi_{ag}^{TM}} \cos\left(\phi_{gs}^{TM}+m\pi\right) \qquad \text{B-(146)}$$

Finally, from Eqs. (88) and (133),

$$A_2 e^{igx} + A_3 e^{-igx} = A_2 e^{-i\phi_{ag}^{TM}}\left[e^{i(gx+\phi_{ag}^{TM})} + e^{-i(gx+\phi_{ag}^{TM})}\right] \qquad \text{B-(147)}$$

$$= 2A_2 e^{-i\phi_{ag}^{TM}} \cos\left(gx+\phi_{ag}^{TM}\right) \qquad \text{B-(148)}$$

It is now possible to write down the TM solutions for $\vec{E}$ and $\vec{H}$ by using Eqs. A-(70) through A-(75) together with Eqs. (87) - (89), (142), (146), and (148)

$$\vec{E}_{TM} = \hat{x}\, E_x + \hat{z}\, E_z \qquad \text{B-(149)}$$

$$\vec{H}_{TM} = \hat{y}\, H_y \qquad \text{B-(150)}$$

$$E_x(x,z,t) = E_{ox}(x)\, e^{i(\omega t - \beta z)} \qquad \text{B-(151)}$$

$$E_z(x,z,t) = E_{oz}(x)\, e^{i(\omega t - \beta z)} \qquad \text{B-(152)}$$

$$H_y(x,z,t) = H_{oy}(x)\, e^{i(\omega t - \beta z)} \qquad \text{B-(153)}$$

where

$$H_{oy}(x) = -2A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{ag}^{TM}\right) e^{-ax} \qquad \text{B-(154)}$$

$$H_{oy}(x) = -2A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(gx + \varphi_{ag}^{TM}\right) \quad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(155)}$$

$$H_{oy}(x) = -2A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{gs}^{TM} + m\pi\right) e^{s(W+x)} \quad \text{for } x \leqq -W \qquad \text{B-(156)}$$

From Eqs. A-(70) and A-(74), $E_x = (i/k_o n^2)\partial H_y/\partial z$ for all x; hence from Eqs. (87) - (89)

$$E_{ox}(x) = \left(\beta/k_o n_a^2\right) H_{oy}(x) \qquad \text{for } x \geqq 0 \qquad \text{B-(157)}$$

$$E_{ox}(x) = \left(\beta/k_o n_g^2\right) H_{oy}(x) \qquad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(158)}$$

$$E_{ox}(x) = \left(\beta/k_o n_s^2\right) H_{oy}(x) \qquad \text{for } x \leqq -W \qquad \text{B-(159)}$$

Last of all, from Eqs. A-(72) and A-(74), $E_z = (-i/k_o n^2)\partial H_y/\partial x$; hence

$$E_{oz}(x) = i\left(\alpha/k_o n_a^2\right) H_{oy}(x) \qquad \text{for } x \geqq 0 \qquad \text{B-(160)}$$

$$E_{oz}(x) = i\left(g/k_o n_g^2\right)\tan\left(gx + \phi_{ag}^{TM}\right) H_{oy}(x) \qquad \text{for } -W \leqq x \leqq 0 \qquad \text{B-(161)}$$

$$E_{oz}(x) = -i\left(s/k_o n_s^2\right) H_{oy}(x) \qquad \text{for } x \leqq -W \qquad \text{B-(162)}$$

APPENDIX C

INHOMOGENEOUS ASYMMETRIC WAVFGUIDES

TE Modes

The following choice of trial solutions is made for F_x (where $\vec{F} = \hat{x} F_x$):

$$F_{xa} = (F_1/i\beta_a) e^{i(\omega t - \beta_a z)} e^{-ax} \qquad \text{for } x \geqq 0 \qquad \text{C-(1)}$$

$$F_{xg} = (1/i\beta_g) e^{i(\omega t - \beta_g z)} \left[F_2 e^{ig} + F_3 e^{-ig}\right] \text{ for } -W \leqq x \leqq 0 \qquad \text{C-(2)}$$

$$F_{xs} = (F_4/i\beta_s) e^{i(\omega t - \beta_s z)} e^{sx} \qquad \text{for } x \leqq -W \qquad \text{C-(3)}$$

This choice is similar to the choice made for the homogeneous case [Fqs. B-(1) through B-(3)], except that now a provision is made in the form of the solution for F_{xg} to allow for an index gradient, i.e., instead of $e^{\pm igx}$ the choice is $e^{\pm ig}$, where

$$g = g(x) \qquad \text{C-(4)}$$

and where it is assumed that all x dependence of F_{xg} is contained within g(x). The function $[g(x)/k_o]$ is often called the eikonal.[5]

From Eqs. A-(99) through A-(104), the choice $\vec{A} = 0$ leads to the following relations for the $\vec{E}$ and $\vec{H}$ fields in the TE case:

$$E_x = 0 \qquad \text{C-(5)}$$

$$E_y = -\frac{\partial F_x}{\partial z} \qquad \text{C-(6)}$$

$$E_z = 0 \qquad \text{C-(7)}$$

$$H_x = (i/k_0)\frac{\partial^2 F_x}{\partial z^2} \qquad \text{C-(8)}$$

$$H_y = 0 \qquad \text{C-(9)}$$

$$H_z = -(i/k_0)\frac{\partial^2 F_x}{\partial z \partial x} \qquad \text{C-(10)}$$

The boundary conditions on $\vec{E}$ and $\vec{H}$ (for no currents in the case of $\vec{H}$) require that the tangential components of $\vec{E}$ and $\vec{H}$ always be continuous across a boundary. This statement is true for any boundary; no distinction is made between a boundary between two different media[6] or a boundary between two different portions of the same medium, including an inhomogeneous medium. The following argument is useful. Eq. A-(7) and A-(8) from Maxwell's equation are integrated over the surface $\tau \equiv \Delta x\, \Delta y$ shown in Figure C-1:

$$\int_\tau \nabla \times \vec{E} \cdot d\vec{\tau} = -\frac{1}{c}\int_\tau \frac{\partial \vec{H}}{\partial t} \cdot d\vec{\tau} \qquad \text{C-(11)}$$

$$\int_\tau \nabla \times \vec{H} \cdot d\vec{\tau} = \frac{1}{c}\int_\tau \frac{\partial (n^2 \vec{E})}{\partial t} \cdot d\vec{\tau} \qquad \text{C-(12)}$$

From Stokes' theorem for any vector $\vec{U}$,

$$\int_\tau \nabla \times \vec{U} \cdot d\vec{\tau} = \oint_\ell \vec{U} \cdot d\vec{\ell} \qquad \text{C-(13)}$$

where ℓ is the contour path surrounding τ, and where the symbol $\oint$ indicates that the integration is over a closed path. Hence Eqs. (11) and (12) become

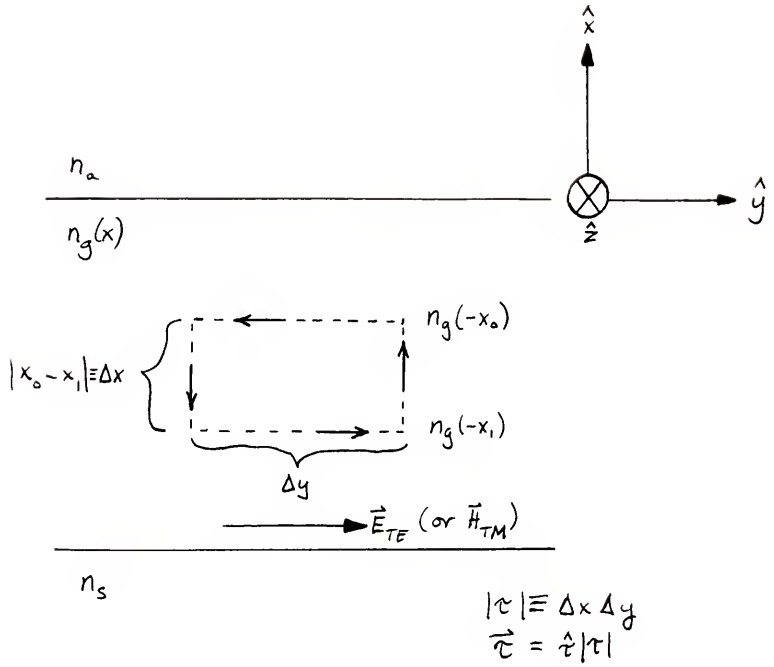


Figure C-1. Rectangular path of integration within an inhomogeneous guide to show behavior of tangential TE (and also TM) component. This view of the guide is one of looking "down the guide" (in the direction of propagation), which is different from that shown in Figure II-1.

$$\oint_{\ell} \vec{E} \cdot d\vec{\ell} = -\frac{1}{C} \int_{\tau} \frac{\partial \vec{H}}{\partial t} \cdot d\vec{\tau} \qquad \text{C-(14)}$$

$$\oint_{\ell} \vec{H} \cdot d\vec{\ell} = \frac{1}{C} \int_{\tau} \frac{\partial (n^2 \vec{E})}{\partial t} \cdot d\vec{\tau} \qquad \text{C-(15)}$$

Referring to Figure C-1, it may be seen that the dot product $\vec{E} \cdot d\vec{\ell}$ will have the following contributions only:

$$\oint_{\ell} \vec{E} \cdot d\vec{\ell} = E^{-}_{TE}(-x_1)\,\Delta y + E^{+}_{TE}(-x_o)(-\Delta y) \qquad \text{C-(16)}$$

where $|F^{-}_{TE}|$ and $|F^{+}_{TE}|$ are the amplitudes of $\vec{F}$ at the levels $x = -x_1$ and $x = -x_o$. A similar equation results for the TM case:

$$\oint_{\ell} \vec{H} \cdot d\vec{\ell} = H^{-}_{TM}(-x_1)\,\Delta y + H^{+}_{TM}(-x_o)(-\Delta y) \qquad \text{C-(17)}$$

which is given here since the relationship for the TM case will be useful in the next section.

Defining the direction of $\vec{\tau}$ as $\hat{\tau} = \vec{\tau}/|\tau|$, then the right-hand sides of Eqs. (14) and (15) can be written as

$$-\frac{1}{C} \int_{\tau} \frac{\partial \vec{H}}{\partial t} \cdot d\vec{\tau} = -\frac{1}{C} \frac{\partial \vec{H}}{\partial t} \cdot \hat{\tau}\ \Delta x \Delta y \qquad \text{C-(18)}$$

$$\frac{1}{C} \int_{\tau} \frac{\partial (n^2 \vec{E})}{\partial t} \cdot d\vec{\tau} = \frac{1}{C} \frac{\partial (n^2 \vec{E})}{\partial t} \cdot \hat{\tau}\ \Delta x \Delta y \qquad \text{C-(19)}$$

Using Eqs. (16) - (19), Eqs. (14) and (15) become

$$\left[E^{-}_{TE}(-x_1) - E^{+}_{TE}(-x_o)\right] \Delta y = -\frac{1}{C} \frac{\partial \vec{H}}{\partial t} \cdot \hat{\tau}\ \Delta x \Delta y \qquad \text{C-(20)}$$

$$\left[H_{TM}^{-}(-x_1) - H_{TM}^{+}(-x_0)\right]\Delta y = \frac{1}{c}\,\frac{\partial(n^2\vec{E})}{\partial t}\,\hat{\tau}\,\Delta x\,\Delta y \qquad \text{C-(21)}$$

Since Δy is arbitrary and is common to both sides of Eqs. (20) and (21), then in the limit as Δx and Δy approach zero, Eqs. (20) and (21) become

$$E_{TE}^{-}(-x_1) - E_{TE}^{+}(-x_0) = -\frac{1}{c}\lim_{\Delta x \to 0}\left[\frac{\partial \vec{H}}{\partial t}\cdot\hat{\tau}\,\Delta x\right] \qquad \text{C-(22)}$$

$$H_{TM}^{-}(-x_1) - H_{TM}^{+}(-x_0) = \frac{1}{c}\lim_{\Delta x \to 0}\left[\frac{\partial(n^2\vec{E})}{\partial t}\cdot\hat{\tau}\,\Delta x\right] \qquad \text{C-(23)}$$

Assuming $\partial\vec{H}/\partial t$ and $\partial(n^2\vec{E}/\partial t$ as well as $\vec{H}$ and $n^2\vec{E}$ remain finite, then the right-hand sides of Eqs. (22) and (23) go to zero as Δx goes to zero. In the limit, then,

$$E_{TE}^{-}(-x_0) - E_{TE}^{+}(-x_0) = 0 \qquad \text{C-(24)}$$

$$H_{TM}^{-}(-x_0) - H_{TM}^{+}(-x_0) = 0 \qquad \text{C-(25)}$$

Since $-x_0$ is an arbitrary position within the guide, then E_{TE} and H_{TM} (which are the only tangential components to the $\hat{x}$-axis) are continuous within an inhomogeneous guide.

Because of this, Eqs. (2) and (24) imply that

$$e^{i\omega t}\left[F_2 e^{ig^{-}(-x_0)} e^{-i\beta_g^{-}(-x_0)z} + F_3 e^{-ig^{-}(-x_0)} e^{+i\beta_g^{-}(-x_0)z}\right]$$

$$= e^{i\omega t}\left[F_2 e^{ig^{+}(-x_0)} e^{-i\beta_g^{+}(-x_0)z} + F_3 e^{-ig^{+}(-x_0)} e^{+i\beta_g^{+}(-x_0)z}\right] \qquad \text{C-(26)}$$

At the level $z = 0$, then the following relationship must exist:

$$g^{-}(x_o) = g^{+}(x_o) \qquad \text{C-(27)}$$

i.e., g(x) is continuous within an inhomogeneous guide.

It is now useful to determine the relationships among β_a, β_g, and β_s using the boundary conditions on $\vec{E}$ and $\vec{H}$. These conditions are given by Eqs. B-(12) through B-(15), and, because of Eqs. (6) and (10), by Eqs. B-(16) and B-(17) too. Substitution of Eqs. (1) - (3) into Eq. (6), together with the boundary conditions given by Eqs. B-(12) and B-(14), produces the following equations:

$$\left(i\beta_a F_{xa}\right)\Big|_{\substack{n_a - n_g \\ \text{boundary}}} = \left(i\beta F_{xg}\right)\Big|_{\substack{n_a - n_g \\ \text{boundary}}} \qquad \text{C-(28)}$$

$$\left(i\beta_s F_{xs}\right)\Big|_{\substack{n_g - n_s \\ \text{boundary}}} = \left(i\beta_g F_{xg}\right)\Big|_{\substack{n_g - n_s \\ \text{boundary}}} \qquad \text{C-(29)}$$

At the n_a - n_g boundary, Eqs. (1), (2), and (28) produce

$$F_1 e^{i(\omega t - \beta_a z)} = e^{i(\omega t - \beta_g z)}\left[F_2 e^{ig(0)} + F_3 e^{-ig(0)}\right] \qquad \text{C-(30)}$$

Since F_1, $F_2\, e^{ig(0)}$, and $F_3\, e^{-ig(0)}$ are constants, then for arbitrary z and t, the following relationship must exist in order that Eq. (30) be satisfied:

$$\beta_a = \beta_g \qquad \text{C-(31)}$$

Similarly, at the n_g - n_s boundary, Eqs. (2), (3), and (29) produce

$$F_4 e^{i(\omega t-\beta_s z)} e^{-sW} = e^{i(\omega t-\beta_g z)}\left[F_2 e^{ig(-W)} + F_3 e^{-ig(-W)}\right] \quad \text{C-(32)}$$

Eq. (32) can be satisfied only if

$$\beta_s = \beta_g \quad \text{C-(33)}$$

Since β_g is the wavevector along the z-axis,

$$\beta_g = k_o\, n_g(x) \sin\Theta(x) \quad \text{C-(34)}$$

But by assumption all the x dependence of F_{xg} is contained within g(x), hence

$$\beta_g(x) = \text{constant} \quad \text{C-(35)}$$

It is convenient, then, from Eqs. (31) and (33)-(35) to let

$$\beta \equiv k_o\, n_g(x)\sin\Theta(x) = \beta_g = \beta_a = \beta_s = \text{constant} \quad \text{C-(36)}$$

$$= k_o \eta \quad \text{C-(37)}$$

where

$$\eta \equiv n_g(x) \sin\theta(x) = \text{constant} \qquad \text{C-(38)}$$

which is simply another way of expressing Snell's law.

Returning now to the wave equation, Eqs. (1) - (3) are substituted into Eq. A-(92), producing

$$\left(-\beta^2 + a^2 + k_o^2 n_a^2\right) F_{xa} = 0 \qquad \text{C-(39)}$$

$$\left[-\beta^2 - (g')^2 + ig'' + k_o^2 n_g^2\right] F_2 e^{ig} + \left[-\beta^2 - (g')^2 - ig'' + k_o^2 n_g^2\right] F_3 e^{-ig} = 0 \qquad \text{C-(40)}$$

$$\left(-\beta^2 + s^2 + k_o^2 n_s^2\right) F_{xs} = 0 \qquad \text{C-(41)}$$

where g' and g'' are the first and second derivatives of g with respect to x. Since either F_2 or F_3 can be set equal to zero, then Eq. (40) can be re-written as two equations:

$$F_2 e^{ig}\left[-\beta^2 - (g')^2 + ig'' + k_o^2 n_g^2\right] = 0 \qquad \text{C-(42)}$$

$$F_3 e^{-ig}\left[-\beta^2 - (g')^2 - ig'' + k_o^2 n_g^2\right] = 0 \qquad \text{C-(43)}$$

The assumption is now made that, for a slowly varying medium (i.e., for a refractive index that varies slowly with x), g'' is small. (This assumption will be discussed in more detail below.) Thus, to a first approximation

$$g'' \approx 0 \qquad \text{C-(44)}$$

For this situation, Eqs. (39) - (43) become [with the aid of Eqs. (36) together with the trigonometric identity $\cos^2\theta = 1 - \sin^2\theta$]

$$a^2 = k_0^2 \left(n_g^2 \sin^2\theta - n_a^2\right)^{1/2} \qquad \text{C-(45)}$$

$$(g')^2 = k_0^2 n_g^2 \cos^2\theta \qquad \text{C-(46)}$$

$$s^2 = k_0^2 \left(n_g^2 \sin^2\theta - n_s^2\right) \qquad \text{C-(47)}$$

Thus, using Eq. (37) in Eqs. (45) - (47) results in the following relations:

$$a = k_0 \left(\eta^2 - n_a^2\right)^{1/2} \qquad \text{C-(48)}$$

$$g' = k_0 q \qquad \text{C-(49)}$$

$$s = k_0 \left(\eta^2 - n_s^2\right)^{1/2} \qquad \text{C-(50)}$$

where

$$q(x) \equiv n_g(x) \cos\theta(x) \qquad \text{C-(51)}$$

Integrating Eq. (49) produces

$$g(x) = k_0 \int_0^x q\, dx \qquad \text{C-(52)}$$

$$= k_0 \int_0^x n_g(x) \cos\theta(x)\, dx \qquad \text{C-(53)}$$

It may be noted from Fq. (53) that, for the homogeneous case ($n_g \cos\theta$ = constant), $g(x) = k_o(n_g \cos\theta)x$, as before; further, $g'' = 0$ is then an exact relation.

Using the value for g' given by Fq. (49), then, to a second-order approximation,

$$g'' = k_o g' \qquad \text{C-(54)}$$

Substitution of Eq. (54) into Eqs. (42) and (43) produces

$$g' = (k_o^2 g^2 \pm i k_o g')^{1/2} \qquad \text{C-(55)}$$

or

$$g' = k_o g (1 \pm i g'/k_o g^2)^{1/2} \qquad \text{C-(56)}$$

where the "+" and "-" refer to the $F_2 e^{ig}$ and the $F_3 e^{-ig}$ terms respectively. Since the right-hand term in the bracket of Fqs. (55) and (56) is small, then g' may be written, to a good approximation as

$$g' \cong k_o g (1 \pm i g'/2 k_o g^2) \qquad \text{C-(57)}$$

or

$$g' = k_o g \pm i g'/2g \qquad \text{C-(58)}$$

hence

$$\frac{dg}{dx} = k_o g \mp \frac{d[\ln(g^{-1/2})]}{dx} \qquad \text{C-(59)}$$

Integrating Eq. (59) produces

$$g(x) = k_o \int_0^x q(x)\,dx \mp i \ln(q^{-1/2}) \qquad \text{C-(60)}$$

Using Eq. (60), the following may be said

$$F_2 e^{ig} + F_3 e^{-ig} = q^{-1/2}\left[F_2 e^{ik_o \int_0^x q\,dx} + F_3 e^{-ik_o \int_0^x q\,dx}\right] \qquad \text{C-(61)}$$

The effect of the second approximation, then, is to cause the amplitude to vary by the factor $q^{-1/2}$. It is now convenient to use Eq. (52) for the value of g(x), writing the imaginary term in Eq. (60) explicitly as an amplitude factor. Thus, $g' = k_o q$ and $g'' = k_o q'$. In this way, the function g(x) will be real and thus will describe the phase only of F_{xg}. From Eq. (?) [and Eq. (36)], then

$$F_{xg} = q^{-1/2}(1/i\beta)\, e^{i(\omega t - \beta z)}\left[F_2 e^{ig} + F_3 e^{-ig}\right] \qquad \text{C-(62)}$$

It is now necessary to determine under what conditions Eq. (62) is a good approximation to the solution of the wave equation. Since, as mentioned previously, e^{ig} and e^{-ig} are orthonormal functions, then each one must independently satisfy the wave equation Eq. A-(84). Taking the first derivative with respect to x of the two terms $e^{\pm ig}$ in F_{xg} leads to the following two relationships:

$$\frac{\partial F_{xg}}{\partial x} = \left(-q'/2q \pm i g'\right)\left[q^{-1/2} F_{2,3}\, e^{\pm ig}\right] e^{i(\omega t - \beta z)} \qquad \text{C-(63)}$$

$$= \left(-q'/2q \pm i k_o q\right)\left[q^{-1/2} F_{2,3}\, e^{\pm ig}\right] e^{i(\omega t - \beta z)} \qquad \text{C-(64)}$$

where, for the sake of brevity, the term $F_{2,3}e^{\pm ig}$ represents the two orthonormal terms F_2e^{ig} and F_3e^{-ig}. The second derivative with respect to x of F_{xg} is given by

$$\frac{\partial^2 F_{xg}}{\partial x^2} = \left\{\left(-g'/2g \pm ik_o g\right)^2 + \left[-g''/2g + (g')^2/2g^2 \pm ik_o g'\right]\right\}\left[g^{-1/2}F_{2,3}\,e^{\pm ig}\right]e^{i(\omega t-\beta z)} \qquad \text{C-(65)}$$

$$= \left\{-k_o^2 g^2 + 3(g')^2/4g^2 - g''/2g\right\}F_{xg} \qquad \text{C-(66)}$$

where the factor in braces in Fq. (66) is valid for both of the terms $F_{2,3}e^{\pm ig}$. Further,

$$\frac{\partial^2 F_{xg}}{\partial z^2} = -\beta^2 F_{xg} = -n_g^2 \sin^2\theta\, F_{xg} \qquad \text{C-(67)}$$

Substituting Eqs. (66) and (67) into Eq. A-(92) produces

$$-\beta^2 - k_o^2 g^2 + 3(g')^2/4g^2 - g''/2g + k_o^2 n_g^2 = 0 \qquad \text{C-(68)}$$

or, since $-\beta^2 + k_o^2 n_g^2 = k_o^2 n_g^2 \cos^2\theta$, then

$$k_o^2 g^2 - \left[3(g')^2/4g^2 - g''/2g\right] = k_o^2 n_g^2 \cos^2\theta \qquad \text{C-(69)}$$

In order that g satisfy the definition in Fq. (51), then in Eq. (69)

$$\left|k_o^2 g^2\right| >> \left|3(g')^2/4g^2 - g''/2g\right| \qquad \text{C-(70)}$$

or

$$\frac{1}{k_o^2}\left|3\left(\frac{q'}{2q^2}\right)^2 - \frac{q''}{2q^3}\right| << 1 \qquad \text{C-(71)}$$

The method described above for solving the wave equation for an inhomogeneous medium is called the WKBJ method,[7-10] and Eq. (71) is the WKBJ approximation for the TE case.[10] The solution given by Eq. (62) [and Eq. (52)], when substituted into Eqs. A-(100), A-(102), and A-(104), is the TE WKBJ-type solution[10] for an inhomogeneous medium. It may be seen that for large wavenumbers k_o (i.e., small λ_o), the WKBJ approximation is more easily satisfied. However, as $q = n_g \cos\theta$ approaches zero, Eq. (71) cannot be satisfied, regardless of how small q' and q" are. This situation will be discussed in Appendix D.

Proceeding now to the boundary conditions, substitution of Eqs. (1), (3), (36), and (62) into Eqs. B-(12) through B-(17) [or, equivalently, Eqs. B-(30) through B-(33)] produces

$$F_1 = q^{-1/2}(0)\left[F_2 e^{ig(0)} + F_3 e^{-ig(0)}\right] \qquad \text{C-(72)}$$

$$-aF_1 = iq^{-1/2}(0)g'(0)\left[F_2 e^{ig(0)} - F_3 e^{-ig(0)}\right] - \frac{i}{2}q^{-3/2}(0)g'(0)\left[F_2 e^{ig(0)} + F_3 e^{-ig(0)}\right] \qquad \text{C-(73)}$$

$$F_4 e^{-sW} = q^{-1/2}(-W)\left[F_2 e^{ig(-W)} + F_3 e^{-ig(-W)}\right] \qquad \text{C-(74)}$$

$$sF_4 e^{-sW} = iq^{-1/2}(-W)g'(-W)\left[F_2 e^{ig(-W)} - F_3 e^{-ig(-W)}\right] - \frac{i}{2}q^{-3/2}(-W)q'(-W)\left[F_2 e^{ig(-W)} + F_3 e^{-ig(-W)}\right] \qquad \text{C-(75)}$$

From Eq. (52),

$$g(0)=0 \quad ; \quad g'(0) = k_0\, g(0) \tag{C-(76)}$$

$$g'(-W) = k_0\, g(-W) \tag{C-(77)}$$

Thus Eqs. (72), (73), and (75) become

$$F_1 = g^{-1/2}(0)\,[F_2 + F_3] \tag{C-(78)}$$

$$-aF_1 = i k_0\, g^{1/2}(0)\,[F_2 - F_3]$$

$$-\frac{i}{2}\, g^{-3/2}(0)\, g'(0)\,[F_2 + F_3] \tag{C-(79)}$$

$$sF_4\, e^{-sW} = i k_0\, g^{1/2}(-W)\left[F_2 e^{ig(-W)} - F_3 e^{-ig(-W)}\right]$$

$$-\frac{i}{2}\, g^{-3/2}(-W)\, g'(0)\left[F_2 e^{ig(-W)} + F_3 e^{-ig(-W)}\right] \tag{C-(80)}$$

The assumption is now made that

$$\left| k_0\, g^{1/2}(x) \right| >> \left| \tfrac{1}{2}\, g^{-3/2}(x)\, g'(x) \right| \tag{C-(81)}$$

or

$$\frac{1}{k_0} \left| \frac{g'}{2g^2} \right| << 1 \tag{C-(82)}$$

which is compatible with the WKBJ approximation in Eq. (71). Using Eq. (82), then Eqs. (79) and (80) become

$$-aF_1 = ik_0 q^{1/2}(0)\left[F_2 - F_3\right] \qquad \text{C-(83)}$$

$$sF_4 e^{-sW} = ik_0 q^{1/2}(-W)\left[F_2 e^{ig(-W)} - F_3 e^{-ig(-W)}\right] \qquad \text{C-(84)}$$

Eqs. (78), (83), (74), and (84) are very similar to Eqs. B-(34) through B-(37); hence these equations can be solved simultaneously in similar fashion. Multiplying Eq. (78) by +a and adding to Eq. (83) results in

$$-F_2\left[aq^{-1/2}(0) + ik_0 q^{1/2}(0)\right] = F_3\left[a\, q^{-1/2}(0) - ik_0 q^{1/2}(0)\right] \qquad \text{C-(85)}$$

or

$$-F_2\left[a + ik_0 q(0)\right] = F_3\left[a - ik_0 q(0)\right] \qquad \text{C-(86)}$$

Multiplying Eq. (74) by -s and adding to Eq. (84) produces

$$-F_2\left[sq^{-1/2}(-W) - ik_0 q^{1/2}(-W)\right] e^{ig(-W)} = F_3\left[sq^{-1/2}(-W) + ik_0 q^{1/2}(-W)\right] e^{-ig(-W)} \qquad \text{C-(87)}$$

or

$$-F_2\left[s - ik_0 q(-W)\right] e^{2ig(-W)} = F_3\left[s + ik_0 q(-W)\right] \qquad \text{C-(88)}$$

Dividing Eq. (86) by (88) results in

$$e^{-2ig(-W)} = \frac{\left[a - ik_0 q(0)\right]}{\left[a + ik_0 q(0)\right]} \quad \frac{\left[s - ik_0 q(-W)\right]}{\left[s + ik_0 q(-W)\right]} \qquad \text{C-(89)}$$

Using Fq. B-(44), Eq. (89) becomes

$$e^{-2ig(-W)} = e^{2i\tan^{-1}[a/k_0 g(0)]}\, e^{2i\tan^{-1}[s/k_0 g(-W)]} \qquad \text{C-(90)}$$

or

$$-2g(-W) - 2\phi_{ag}^{TE} - 2\phi_{gs}^{TE} = 2m\pi \qquad \text{C-(91)}$$

where

$$\phi_{ag}^{TE} \equiv \tan^{-1}\left[a/k_0 g(0)\right] \qquad \text{C-(92)}$$

$$\phi_{gs}^{TE} \equiv \tan^{-1}\left[s/k_0 g(-W)\right] \qquad \text{C-(93)}$$

and, from Eqs. (48), (50), and (51)

$$\frac{a}{k_0 g(0)} = \frac{(\eta^2 - n_a^2)^{1/2}}{n_g(0)\cos\theta(0)} \qquad \text{C-(94)}$$

$$\frac{s}{k_0 g(-W)} = \frac{(\eta^2 - n_s^2)^{1/2}}{n_g(-W)\cos\theta(-W)} \qquad \text{C-(95)}$$

Using Fq. (53), Fq. (91) becomes the waveguide equation:

$$2k_0 \int_{-W}^{0} n_g(x)\cos\theta(x)\,dx - 2\phi_{ag}^{TE} - 2\phi_{gs}^{TE} = 2m\pi \qquad \text{C-(96)}$$

It may be noted that when n_g (and thus cos θ too) is constant within the guide, then Eq. (96) reduces to the waveguide equation for the homogeneous case, Fq. B-(52). Using Eq. (38) the integrand in Fq. (96) can be written as

$$q(x) = n_g(x)\cos\theta(x) = \left[n_g^2(x) - \eta^2\right]^{1/2} \qquad \text{C-(97)}$$

Also,

$$\frac{a}{k_0 q(0)} = \frac{(\eta^2 - n_a^2)^{1/2}}{\left[n_g^2(0) - \eta^2\right]^{1/2}} \qquad \text{C-(98)}$$

$$\frac{s}{k_0 q(-W)} = \frac{\eta^2 - n_s^2}{\left[n_g^2(-W) - \eta^2\right]^{1/2}} \qquad \text{C-(99)}$$

Thus, Eq. (96) becomes

$$2k_0 \int_{-W}^{0} \left[n_g^2(x) - \eta^2\right]^{1/2} - 2\tan^{-1}\left\{\frac{(\eta^2 - n_a^2)^{1/2}}{\left[n_g^2(0) - \eta^2\right]^{1/2}}\right\} - 2\tan^{-1}\left\{\frac{\eta^2 - n_s^2}{\left[n_g^2(-W) - \eta^2\right]^{1/2}}\right\} = 2m\pi \qquad \text{C-(100)}$$

It may be seen that Eq. (100) is transcendental in both η and W. Writing

$$\eta = n_g(0)\sin\theta(0) \qquad \text{C-(101)}$$

then, for a given value of $n_g(0)$, perhaps the easiest way to solve Eq. (100) is to pick a value for $\theta(0)$, then try values of W until one is found which satisfies Eq. (100). By using different values of $\theta(0)$, a set of curves can be generated for the different modes m.

Eq. (96) [or Eq. (100)] can be used to obtain the cutoff thickness $(W_{min})_{TE}$ of a guide. There are two possibilities for waveguide cutoff. The critical angle at the n_g - n_s boundary is defined by

$$\theta_{cgs} \equiv \sin^{-1}\left[n_a / n_{gs}\right] \qquad \text{C-(102)}$$

while the critical angle at the n_a - n_g boundary is given by

$$\theta_{cag} \equiv \sin^{-1}\left[n_a / n_{ag} \right] \qquad \text{C-(103)}$$

where

$$n_{gs} \equiv n_g(-w) \qquad \text{C-(104)}$$

$$n_{ag} \equiv n_g(0) \qquad \text{C-(105)}$$

Now, if

$$\theta(-w) \leqq \theta_{cgs} \qquad \text{C-(106)}$$

or

$$\theta(0) \leqq \theta_{cag} \qquad \text{C-(107)}$$

then the light waves will escape or leak out. From Eqs. (102) - (105), the cutoff conditions become

$$\sin\theta(-w) \leqq \frac{n_s}{n_{gs}} \qquad \text{C-(108)}$$

$$\sin\theta(0) \leqq \frac{n_a}{n_{ag}} \qquad \text{C-(109)}$$

or

$$n_g(-w)\sin\theta(-w) \leqq n_s \qquad \text{C-(110)}$$

$$n_g(0)\sin\theta(0) \leqq n_a \qquad \text{C-(111)}$$

But, by Eq. (38), Eqs. (110) and (111) become

$$\eta = n_g(x)\sin\theta(x) \leqq n_s \qquad \text{C-(112)}$$

$$\eta = n_g(x)\sin\theta(x) \leqq n_a \qquad \text{C-(113)}$$

It was assumed for homogeneous waveguides that $n_s > n_a$. Hence as θ decreases from a maximum value of $\pi/2$ (propagation straight down the guide), then, for a given functional form of n_g, the cutoff condition first to be met is given by Eq. (112). That is, waves will start to leak out at the $n_g - n_s$ boundary when $\sin\theta(x) = n_s/n_g(x)$: they will still be totally internally reflected at the $n_a - n_g$ boundary. This result is identical with the result obtained for the homogeneous case. Substituting Eq. (112) into Eqs. (92) - (95) and Eqs. (98) - (99) produces $\phi_{gs}^{TE} = 0$ and

$$\rho_{ag}^{TE} = \tan^{-1}\left[\left(n_s^2 - n_a^2\right)^{1/2} \Big/ \left(n_{ag}^2 - n_a^2\right)^{1/2}\right] \qquad \text{C-(114)}$$

Hence at cutoff, Eq. (100) becomes the following:

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_s^2\right]^{1/2} dx = \frac{\lambda_o}{2\pi}\left\{m\pi + \tan^{-1}\left[\frac{\left(n_s^2 - n_a^2\right)^{1/2}}{\left(n_{ag}^2 - n_s^2\right)^{1/2}}\right]\right\} \qquad \text{C-(115)}$$

For the inhomogeneous case, the situation $n_s < n_a$ is different than $n_s > n_a$ (since $n_{ag} \neq n_{gs}$). Hence if $n_s < n_a$, then the cutoff condition is given by Eq. (113), in which case $\phi_{gs}^{TE} \neq 0$, $\phi_{ag}^{TE} = 0$, and Eq. (100) becomes

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \frac{\lambda_o}{2\pi}\left\{m\pi + \tan^{-1}\left[\frac{\left(n_a^2 - n_s^2\right)^{1/2}}{\left(n_{gs}^2 - n_a^2\right)^{1/2}}\right]\right\} \qquad \text{C-(116)}$$

In order to find the final solutions for $\vec{E}$ and $\vec{H}$, it is necessary to calculate the coefficients F_1, F_2, F_3, and F_4, where, as before, F_2 is chosen as the arbitrary amplitude coefficient. Letting

$$F_3 = e^{i2\phi^{TE}} F_2 \qquad \text{C-(117)}$$

Then Eqs. (86) and (88) become

$$-\left[a + ik_0 g(0)\right] = e^{i2\phi^{TE}} \left[a - ik_0 g(0)\right] \qquad \text{C-(118)}$$

$$-\left[s - ik_0 g(-w)\right] e^{2ig(-w)} = e^{i2\phi^{TE}} \left[s + ik_0 g(-w)\right] \qquad \text{C-(119)}$$

Dividing the complex conjugate of Eq. (119) into (118) (and re-arranging factors) produces

$$e^{i4\phi^{TE}} = \frac{a + ik_0 g(0)}{a - ik_0 g(0)} \; \frac{s - ik_0 g(-w)}{s + ik_0 g(-w)} \; e^{2ig(-w)} \qquad \text{C-(120)}$$

Using Eqs. B-(44), (92), and (93), then

$$e^{i4\phi^{TE}} = e^{-2i\phi_{ag}^{TE}} \; e^{+2i\phi_{gs}^{TE}} \; e^{2ig(-w)} \qquad \text{C-(121)}$$

or

$$4\phi^{TE} = -2\phi_{ag}^{TE} + 2\phi_{gs}^{TE} + 2g(-w) + 2m\pi \qquad \text{C-(122)}$$

Using Eq. (91), Eq. (122) implies that

$$\phi^{TE} = -\phi_{ag}^{TE} \qquad \text{C-(123)}$$

Turning now to F_1 and F_4, Eqs. (78) and (74) can be solved in terms of F_2. Using Eqs. (117) and (123), Eq. (78) becomes

$$F_1 = g^{-1/2}(0)\ F_2\left(1 + e^{-i2\phi_{ag}^{T\bar{E}}}\right) \qquad \text{C-(124)}$$

$$= g^{-1/2}(0)\ F_2\ e^{-i\phi_{ag}^{T\bar{E}}}\ 2\cos\left(\phi_{ag}^{T\bar{E}}\right) \qquad \text{C-(125)}$$

Also, from Eq. (74),

$$F_4 = e^{sW}\, g^{-1/2}(-W)\, F_2\left[e^{ig(-W)} + e^{-i2\phi_{ag}^{T\bar{E}}}\, e^{-ig(-W)}\right] \qquad \text{C-(126)}$$

$$= e^{sW}\, g^{-1/2}(-W)\, e^{-i\phi_{ag}^{T\bar{E}}}\left\{e^{i\left[g(-W)+\phi_{ag}^{T\bar{E}}\right]} + e^{-i\left[g(-W)+\phi_{ag}^{T\bar{E}}\right]}\right\} F_2 \qquad \text{C-(127)}$$

or, from Eq. (91)

$$F_4 = e^{sW}\, g^{-1/2}(-W)\, F_2\, e^{-i\phi_{ag}^{T\bar{E}}}\left[e^{-i\left(\phi_{gs}^{T\bar{E}} + m\pi\right)} + e^{i\left(\phi_{gs}^{T\bar{E}} + m\pi\right)}\right] \qquad \text{C-(128)}$$

$$= 2F_2\, e^{sW}\, e^{-i\phi_{gs}^{T\bar{E}}}\ \cos\left(\phi_{gs}^{T\bar{E}} + m\pi\right) \qquad \text{C-(129)}$$

Finally, from Eqs. (62), (117), and (123)

$$g^{-1/2}\left[F_2 e^{ig} + F_3 e^{-ig}\right] = g^{-1/2}\, F_2\, e^{-i\phi_{ag}^{T\bar{E}}}\left[e^{i\left(g+\phi_{ag}^{T\bar{E}}\right)} + e^{-i\left(g+\phi_{ag}^{T\bar{E}}\right)}\right] \qquad \text{C-(130)}$$

$$= 2g^{-1/2}\, F_2\, e^{-i\phi_{ag}^{T\bar{E}}}\ \cos\left[g(x) + \phi_{ag}^{T\bar{E}}\right] \qquad \text{C-(131)}$$

It may be noted that, except for the factors $q^{-1/2}(0)$ and $q^{-1/2}(-W)$, Eqs. (125) and (127) are identical in form with Eqs. B-(68) and B-(72) providing, of course, that the appropriate values of ϕ_{ag}^{TE} and ϕ_{gs}^{TE} are used. Eq. (131) is similar to Eq. B-(74), with the exceptions that the term gx is replaced by g(x) for the inhomogeneous case: also, there is the additional factor $q^{-1/2}(x)$. Further, since Eqs. (5) - (10) are identical to Fqs. A-(59) through A-(64) for the homogeneous case, it is possible to write down the solutions for $\vec{E}_{TE}$ and $\vec{H}_{TE}$ for the inhomogeneous case by using Eqs. B-(75) through B-(86), providing of course that the term gx is replaced by g(x), and also that the factor (g/k_o) in Eq. B-(85) is replaced by $[g'(x)/k_o] = q(x)$. [There is an additional term for H_{oz} due to the x derivative of $q^{-1/2}(x)$, as given by Eq. (10). However, according to Eq. (81), this term is negligible in the WKBJ approximation.] Thus, E_{ov}, H_{ox}, and H_{oz} for the inhomogeneous are

$$E_{oy}(x) = 2q^{-1/2}(0)\, F_2 e^{-i\phi_{ag}^{TE}} \cos\left(\phi_{ag}^{TE}\right) e^{-ax} \quad \text{for } x \geqq 0 \qquad \text{C-(132)}$$

$$E_{oy}(x) = 2q^{-1/2}(x)\, F_2 e^{-i\phi_{ag}^{TE}} \cos\left[g(x) + \phi_{ag}^{TE}\right] \quad \text{for } -W \leqq x \leqq 0 \qquad \text{C-(133)}$$

$$E_{oy}(x) = 2q^{-1/2}(-W)\, F_2 e^{-i\phi_{ag}^{TE}} \cos\left(\phi_{gs}^{TE} + m\pi\right) e^{s(W+x)} \quad \text{for } x \leqq -W \qquad \text{C-(134)}$$

$$H_{ox}(x) = -(\beta/k_o) E_{oy}(x) \quad \text{for all } x \qquad \text{C-(135)}$$

$$H_{oz}(x) = -i(a/k_o) E_{oy}(x) \quad \text{for } x \geqq 0 \qquad \text{C-(136)}$$

$$H_{oz}(x) = -i\, q(x) \tan\left[g(x) + \phi_{ag}^{TE}\right] E_{oy}(x) \quad \text{for } -W \leqq x \leqq 0 \qquad \text{C-(137)}$$

$$H_{oz}(x) = +i(s/k_2) E_{oy}(x) \quad \text{for } x \leqq -W \qquad \text{C-(138)}$$

TM Modes

The following choice of trial solutions is made for A_x (where $\vec{A} = \hat{x} A_x$)

$$A_{xa} = \left(A_1 / i\beta_a\right) e^{i(\omega t - \beta_a z)} e^{-ax} \quad \text{for} \quad x \geq 0 \qquad \text{C-(139)}$$

$$A_{xg} = \left(1/i\beta_g\right) e^{i(\omega t - \beta_g z)} \left[A_2 e^{ig} + A_3 e^{-ig}\right] \quad \text{for} \quad -W \leq x \leq 0 \qquad \text{C-(140)}$$

$$A_{xs} = \left(A_4 / i\beta_s\right) e^{i(\omega t - \beta_s z)} e^{sx} \quad \text{for} \quad x \leq -W \qquad \text{C-(141)}$$

From Eqs. A-(99) through A-(104), the choice $\vec{F} = 0$ leads to the following relations for the $\vec{E}$ and $\vec{H}$ fields in the TM case:

$$E_x = \left[i/k_0 n^2(x)\right] \frac{\partial^2 A_x}{\partial z^2} \qquad \text{C-(142)}$$

$$E_y = 0 \qquad \text{C-(143)}$$

$$E_z = -\left[i/k_0 n^2(x)\right] \frac{\partial^2 A_x}{\partial z \partial x} \qquad \text{C-(144)}$$

$$H_x = 0 \qquad \text{C-(145)}$$

$$H_y = \frac{\partial A_x}{\partial z} \qquad \text{C-(146)}$$

$$H_z = 0 \qquad \text{C-(147)}$$

It is useful to first determine the relationships among β_a, β_g, and β_s using the boundary conditions on $\vec{E}$ and $\vec{H}$ given by Fqs. B-(93)

through B-(96). Using Eq. (146), then

$$\left(i\beta_a A_{xa}\right)\Big|_{\substack{n_a-n_g \\ \text{boundary}}} = \left(i\beta_g A_{xg}\right)\Big|_{\substack{n_a-n_g \\ \text{boundary}}} \qquad \text{C-(148)}$$

$$\left(i\beta_s A_{xs}\right)\Big|_{\substack{n_g-n_s \\ \text{boundary}}} = \left(i\beta_g A_{xg}\right)\Big|_{\substack{n_g-n_s \\ \text{boundary}}} \qquad \text{C-(149)}$$

Using Eqs. (139) - (140), Eq. (148) and (149) become

$$A_1 e^{i(\omega t-\beta_a z)} = e^{i(\omega t-\beta_g z)}\left[A_2 e^{ig(0)} + A_3 e^{-ig(0)}\right] \qquad \text{C-(150)}$$

$$A_4 e^{i(\omega t-\beta_s z)} e^{-sW} = e^{-i(\omega t-\beta_g z)}\left[A_2 e^{ig(-W)} + A_3 e^{-ig(-W)}\right] \qquad \text{C-(151)}$$

Eqs. (150) and (151) can be satisfied for arbitrary z only if

$$\beta_a = \beta_g \qquad \text{C-(152)}$$

$$\beta_s = \beta_g \qquad \text{C-(153)}$$

Because H_{TM} (and therefore A_x) must be continuous within the guide, even if inhomogeneous [as given by Eq. (32) for two arbitrary adjacent points x_0 and x_1 within the guide], then β_g must be a constant within the guide. It is convenient once again then to let β be defined by Eq. (36).

Substituting Eqs. (139) - (141) into the wave equation Eq. A-(93) produces

$$\left(-\beta^2 + a^2 + k_o^2 n_a^2\right) F_{xa} = 0 \qquad \text{C-(154)}$$

$$\left[-\beta^2-(g')^2+ig''-\frac{1}{n_g^2}ig'(n_g^2)'+k_o^2n_g^2\right]F_2e^{ig}$$
$$+\left[-\beta^2-(g')^2-ig''+\frac{1}{n_g^2}ig'(n_g^2)'+k_o^2n_g^2\right]F_3e^{-ig}=0 \qquad \text{C-(155)}$$

$$\left(-\beta^2+s^2+k_o^2n_s^2\right)F_{xs}=0 \qquad \text{C-(156)}$$

Solving Eqs. (154) and (156) leads to the results given in Fqs. (48) and (50). Again, since e^{ig} and e^{-ig} are orthonormal functions, Eq. C-(155) can be re-written as two equations:

$$-\beta^2-(g')^2+ig''-\frac{2ig'n_g'}{n_g}+k_o^2n_g^2=0 \qquad \text{C-(157)}$$

$$-\beta^2-(g')^2-ig''+\frac{2ig'n_g'}{n_g}+k_o^2n_g^2=0 \qquad \text{C-(158)}$$

To a first approximation

$$g''-\frac{2g'n_g'}{n_g}\approx 0 \qquad \text{C-(159)}$$

When this is done, both Eqs. (157) and (158) have the result given by Eqs. (49) and (52) for g'(x) and g(x). Using Fq. (49), then, to a second approximation

$$g''-\frac{2g'n_g'}{n_g}=k_og'-\frac{2k_ogn_g'}{n_g} \qquad \text{C-(160)}$$

and then from Eqs. (157) and (158)

$$g'=\left[k_o^2g^2\pm i\left(k_og'-\frac{2k_ogn_g'}{n_g}\right)\right]^{1/2} \qquad \text{C-(161)}$$

$$= k_0 q\left[1 \pm \frac{i}{k_0 q}\left(\frac{q'}{q} - 2\frac{n_g'}{n_g}\right)\right]^{1/2} \qquad \text{C-(162)}$$

$$\cong k_0 q\left[1 \pm \frac{1}{2}\frac{i}{k_0 q}\left(\frac{q'}{q} - 2\frac{n_g'}{n_g}\right)\right] \qquad \text{C-(163)}$$

$$= k_0 q \mp i\left\{\frac{d\left[\ln(q^{-1/2})\right]}{dx} + \frac{d(\ln n_g)}{dx}\right\} \qquad \text{C-(164)}$$

hence

$$g(x) = k_0 \int_0^x q\,dx \mp i\left(\ln q^{-1/2} + \ln n_g\right) \qquad \text{C-(165)}$$

Thus,

$$A_2 e^{ig} + A_3 e^{-ig} = n_g q^{-1/2}\left[A_2 e^{ik_0 \int_0^x q\,dx} + A_3 e^{-ik_0 \int_0^x q\,dx}\right] \qquad \text{C-(166)}$$

which is a result similar to Eq. (61) for the $\vec{E}$ vector in the TE case except for the additional factor n_g. Again it is convenient to use Eq. (52) for the value of g(x), letting the imaginary part in Eq. (165) be explicitly written as an amplitude factor. Thus

$$A_{xg} = n_g q^{-1/2}(1/i\beta) e^{i(\omega t - \beta z)}\left[A_2 e^{ig} + A_3 e^{-ig}\right] \qquad \text{C-(167)}$$

It is now necessary to determine under what conditions Eq. (167) is a good approximation to the solution of the wave equation. Since either A_2 or A_3 can be set equal to zero, then the two terms $A_2 e^{ig}$ and $A_3 e^{-ig}$ must independently satisfy the wave equation. Taking the first derivative with respect to x leads to the following two relationships

$$\frac{\partial A_{xg}}{\partial x} = \left(n_g'/n_g - q'/2q \pm i k_0 q\right)\left[n_g q^{-1/2} A_{2,3}\, e^{\pm ig}\right] e^{i(\omega t - \beta z)} \qquad \text{C-(168)}$$

where, for the sake of brevity, the term $A_{2,3}e^{\pm ig}$ represents A_2e^{+ig} and A_3e^{-ig}. The second derivative is

$$\frac{\partial^2 A_{xg}}{\partial x^2} = \left\{\left(n_g'/n_g - g'/2g \pm ik_0 g\right)^2 + \left[n_g''/n_g - \left(n_g'\right)^2/n_g^2 - g''/2g + \left(g'\right)^2/2g^2 \pm ik_0 g'\right]\right\}\left[n_g g^{-1/2} A_{2,3} e^{\pm ig}\right] \qquad \text{C-(169)}$$

Also, the third term in the wave equation, Eq. A-(93), is

$$-\frac{1}{n_g^2}\frac{\partial A_{xg}}{\partial x}\frac{\partial n_g^2}{\partial x} = -\frac{2n_g'}{n_g}\frac{\partial A_{xg}}{\partial x} \qquad \text{C-(170)}$$

Adding Eqs. (169) and (170) results in

$$\frac{\partial^2 A_{xg}}{\partial x^2} - \frac{2n_g'}{n_g}\frac{\partial A_{xg}}{\partial x} = \left\{-k_0 g^2 + \frac{3}{4}\left(\frac{g'}{g}\right)^2 - \frac{g''}{2g} - 2\left(\frac{n_g'}{n_g}\right)^2 + \frac{n_g''}{n_g}\right\} A_{xg} \qquad \text{C-(171)}$$

where the factor in braces in Eq. (171) is valid for both of the terms $A_{2,3}e^{\pm ig}$. Further

$$\frac{\partial^2 A_{xg}}{\partial z^2} = -\beta^2 A_{xg} = -n_g^2 \sin^2\theta\, A_{xg} \qquad \text{C-(172)}$$

Substituting Eqs. (171) and (172) into Eq. A-(93) produces

$$+k_0 g^2 - \left[\frac{3}{4}\left(\frac{g'}{g}\right)^2 - \frac{g''}{2g} - 2\left(\frac{n_g'}{n_g}\right)^2 + \frac{n_g''}{n_g}\right] = k_0^2 n_g^2 \cos^2\theta \qquad \text{C-(173)}$$

In order that q satisfy the definition in Eq. (51), then in Eq. (173)

$$\left| k_o^2 q^2 \right| >> \left| \frac{3}{4}\left(\frac{q'}{q}\right)^2 - \frac{q''}{2q} - 2\left(\frac{n_g'}{n_g}\right)^2 + \frac{n_g''}{n_g} \right| \qquad \text{C-(174)}$$

or

$$\frac{1}{k_o^2}\left| 3\left(\frac{q'}{2q^2}\right)^2 - \frac{q''}{2q^3} + \frac{1}{q^2}\left\{ \frac{n_g''}{n_g} - 2\left(\frac{n_g'}{n_g}\right)^2 \right\} \right| << 1 \qquad \text{C-(175)}$$

This is the WKBJ approximation for an inhomogeneous medium for the TM case.[10] It has the two additional terms with n_g' and n_g'' that TE approximation, Eq. (71), does not have. The solution given by Eq. (167) [and Eq. (52)] is the TM WKBJ-type solution.[10] As q approaches zero, Eq. (175) cannot be satisfied, and this situation will be discussed in Appendix D.

Proceeding now to the boundary conditions, substitution of Eqs. (139), (141), (36), and (167) into Eqs. B-(90) through B-(98) [or, equivalently, Fqs. B-(110) through B-(113)] produces

$$-n_a^{-2} a A_1 = i n_g^{-1}(0) q^{-1/2}(0) g'(0)\left[A_2 e^{ig(0)} - A_3 e^{-ig(0)} \right] + n_g^{-2}(0)\left[n_g'(0) q^{-1/2}(0) - n_g(0) g'(0)/2q^{3/2}(0) \right]\left[A_2 e^{ig(0)} - A_3 e^{-ig(0)} \right] \qquad \text{C-(176)}$$

$$A_1 = n_g(0)\, q^{-1/2}(0)\left[A_2 e^{ig(0)} + A_3 e^{-ig(0)} \right] \qquad \text{C-(177)}$$

$$+n_s^{-2} s A_4 e^{-sW} = i n_g^{-1}(-W) q^{-1/2}(-W) g'(-W) \left[A_2 e^{ig(-W)} - A_3 e^{-ig(-W)}\right]$$
$$+ n_g^{-2}(-W)\left[n_g'(-W) q^{-1/2}(-W) - n_g(-W) q'(-W)/2 q^{3/2}(-W)\right]\left[A_2 e^{ig(-W} + A_3 e^{-ig(-W)}\right]$$

C-(178)

$$A_4 e^{-sW} = n_g(-W) q^{-1/2}(-W)\left[A_2 e^{ig(-W)} + A_3 e^{-ig(-W)}\right] \qquad \text{C-(179)}$$

The values of g(0), g'(0), g'(-W) are given by Eqs. (75) - (77). Also, the assumption is made that

$$\left| n_g^{-1}(x) q^{-1/2}(x) g'(x) \right| >> \left| \frac{n_g'(x)}{n_g^2(x) q^{1/2}(x)} - \frac{q'(x)}{2 n_g(x) q^{3/2}(x)} \right| \qquad \text{C-(180)}$$

Since $g'(x) = k_o q(x)$, then Eq. (180) becomes

$$\left| \frac{k_o q^{1/2}}{n_g} \right| >> \left| \frac{n_g'}{n_g^2 q^{1/2}} - \frac{q'}{2 n_g q^{3/2}} \right| \qquad \text{C-(181)}$$

or

$$\frac{1}{k_o} \left| \frac{n_g'}{n_g q} - \frac{q'}{2 q^2} \right| << 1 \qquad \text{C-(182)}$$

which is the same order of magnitude (or better) approximation as the WKBJ approximation given by Eq. (175). Using Eq. (181) together with Eqs. (75) - (77), Eqs. (176) and (178) can be greatly simplified:

$$- n_a^{-2} a A_1 = i n_g^{-1}(0) k_o q^{1/2}(0) \left[A_2 - A_3\right] \qquad \text{C-(183)}$$

$$+ n_s^{-2} s A_4 e^{-sW} = i n_g^{-1}(-W) k_o q^{1/2}(-W)\left[A_2 e^{ig(-W)} - A_3 e^{-ig(-W)}\right] \qquad \text{C-(184)}$$

Using Eq. (75), Eq. (177) becomes

$$A_i = n_g(0)\, g^{-1/2}(0)\, [A_2 + A_3] \qquad \text{C-(185)}$$

Multiplying Eq. (185) by $+an_a^{-2}$ and adding it to Eq. (183) produces:

$$-A_2 \left[a n_a^{-2}\, n_g(0)\, g^{-1/2}(0) + i n_g^{-1}(0)\, k_0\, g^{1/2}(0) \right] = A_3 \left[a n_a^{-2} n_g(0)\, g^{-1/2}(0) - i n_g^{-1}(0) k_0 g^{1/2}(0) \right] \qquad \text{C-(186)}$$

or

$$-A_2 \left[a\, n_g^2(0)/n_a^2 + i\, k_0\, g(0) \right] = A_3 \left[a\, n_g^2(0)/n_a^2 - i k_0\, g(0) \right] \qquad \text{C-(187)}$$

Multiplying Eq. (179) by $-sn_s^{-2}$ and adding it to Eq. (184) produces

$$-A_2 \left[-s n_s^{-2}\, n_g(-w)\, g^{-1/2}(-w) + i\, n_g^{-1}(-w) k_0\, g^{1/2}(-w) \right] = A_3 \left[-s n_s^{-2} n_g(-w)\, g^{-1/2}(-w) - i n_g^{-1}(-w) k_0\, g^{1/2}(-w) \right] \qquad \text{C-(188)}$$

or

$$-A_2 e^{ig(-w)} \left[s n_g^2(-w)/n_s^2 - i k_0\, g(-w) \right] = A_3 e^{-ig(-w)} \left[s n_g^2(-w)/n_s^2 + i k_0\, g(-w) \right] \qquad \text{C-(189)}$$

Dividing Eq. (187) by Eq. (189) results in

$$\frac{a\, n_g^2(0)/n_a^2 + i k_0\, g(0)}{s\, n_g^2(-w)/n_s^2 - i\, k_0\, g(-w)}\, e^{-ig(-w)} = \frac{a n_g^2(0)/n_a^2 - i k_0\, g(0)}{s\, n_g^2(-w)/n_s^2 + i k_0 g(-w)}\, e^{ig(-w)} \qquad \text{C-(190)}$$

or

$$e^{-2ig(-w)} = \frac{a n_g^2(0)/n_a^2 - ik_0 g(0)}{a n_g^2(0)/n_a^2 + ik_0 g(0)} \quad \frac{s n_g^2(-w)/n_s^2 - ik_0 g(-w)}{s n_g^2(-w)/n_s^2 + ik_0 g(-w)} \qquad \text{C-(191)}$$

Using Fq. B-(44), Eq. (89) becomes

$$\exp\left[-2ig(-w)\right] = \exp\left\{2i\tan^{-1}\left[a n_g^2(0)/n_a^2 k_0 g(0)\right] + 2i\tan^{-1}\left[s n_g^2(-w)/n_s^2 k_0 g(-w)\right]\right\} \qquad \text{C-(192)}$$

or

$$-2g(-w) - 2\phi_{ag}^{TM} - 2\phi_{gs}^{TM} = 2m\pi \qquad \text{C-(193)}$$

where

$$\phi_{ag}^{TM} \equiv \tan^{-1}\left[a n_g^2(0)/n_a^2 k_0 g(0)\right] \qquad \text{C-(194)}$$

$$\phi_{gs}^{TM} \equiv \tan^{-1}\left[s n_g^2(-w)/n_s^2 k_0 g(-w)\right] \qquad \text{C-(195)}$$

Using Eqs. (48), (50), and (51), then the arguments of the arctangents in Eqs. (194) and (195) become

$$\frac{a n_g^2(0)}{n_a^2 k_0 g(0)} = \frac{n_g^2(0)}{n_a^2} \quad \frac{(\eta^2 - n_a^2)^{1/2}}{n_g(0)\cos\theta(0)} \qquad \text{C-(196)}$$

$$\frac{s n_g^2(-w)}{n_s^2 k_0 g(-w)} = \frac{n_g^2(-w)}{n_s^2} \quad \frac{(\eta^2 - n_s^2)^{1/2}}{n_g(-w)\cos\theta(-w)} \qquad \text{C-(197)}$$

It may be noted that ϕ_{ag}^{TM} and ϕ_{gs}^{TM} are identical to ϕ_{ag}^{TE} and ϕ_{gs}^{TE} in Eqs. (92) - (95) except for the factors $n_g^2(0)/n_a^2$ and $n_g^2(-W)/n_a^2$. Using Eq. (53), Eq. (193) becomes

$$2k_o \int_{-W}^{0} n_g(x) \cos\theta(x)\, dx - 2\varphi_{ag}^{TM} - 2\varphi_{gs}^{TM} = 2m\pi \qquad \text{C-(198)}$$

which is the waveguide equation for the TM case. Again it may be noted that when n_g and cos θ are constant within the guide, then Eq. (198) reduces to the waveguide equation for the homogeneous case, Eq. B-(129).

Using Eqs. (97), (194)-(197), Eq. (198) becomes

$$2k_o \int_{-W}^{0} \left[n_g^2(x) - \eta^2 \right]^{1/2} dx - 2\tan^{-1}\left\{ \frac{n_g^2(0)}{n_a^2} \frac{(\eta^2 - n_a^2)^{1/2}}{[n_g^2(0) - \eta^2]^{1/2}} \right\}$$

$$-2\tan^{-1}\left\{ \frac{n_g^2(-W)}{n_s^2} \frac{(\eta^2 - n_s^2)^{1/2}}{[n_g^2(-W) - \eta^2]^{1/2}} \right\} = 2m\pi \qquad \text{C-(199)}$$

It may be seen that Eq. (199) is transcendental in η and W, similar to Eq. (100).

Following the arguments preceding Eq. (114), the cutoff condition is given by Eq. (112) or Eq. (113). Hence at cutoff for $n_s > n_a$, Eq. (199) becomes

$$\int_{-(W_{min})_{TM}}^{0} \left[n_g^2(x) - n_s^2 \right]^{1/2} dx = \frac{\lambda_o}{2\pi} \left\{ m\pi + \tan^{-1}\left[\frac{n_{ag}^2}{n_a^2} \frac{(n_s^2 - n_a^2)^{1/2}}{(n_{ag}^2 - n_s^2)^{1/2}} \right] \right\} \qquad \text{C-(200)}$$

For $n_s < n_a$, Eq. (199) becomes at cutoff

$$\int_{-(W_{min})_{TM}}^{0} \left[n_g^2(x) - n_a^2 \right]^{1/2} dx = \frac{\lambda_o}{2\pi} \left\{ m\pi + \tan^{-1}\left[\frac{n_{gs}^2}{n_s^2} \frac{(n_a^2 - n_s^2)^{1/2}}{(n_{gs}^2 - n_a^2)^{1/2}} \right] \right\} \qquad \text{C-(201)}$$

Eqs. (200) and (201) are similar to Eqs. (115) and (116) for the TE case.

In order to find the final solutions for $\vec{E}$ and $\vec{H}$ for the TM case, it is necessary to solve for A_1, A_3, and A_4 in terms of the arbitrary amplitude A_2. Letting

$$A_3 = e^{i2\varphi^{TM}} A_2 \qquad \text{C-(202)}$$

then Eqs. (187) and (189) become

$$-\left[a n_{ag}^2/n_a^2 + i k_0 g(0)\right] = e^{i2\varphi^{TM}}\left[a n_{ag}^2/n_a^2 - i k_0 g(0)\right] \qquad \text{C-(203)}$$

$$-e^{ig(-w)}\left[s n_{gs}^2/n_s^2 - i k_0 g(-w)\right] = e^{i2\varphi^{TM}} e^{-ig(-w)}\left[s n_{gs}^2/n_s^2 + i k_0 g(-w)\right] \qquad \text{C-(204)}$$

Taking the complex conjugate of Eq. (204) and dividing it into Eq. (203) produces

$$e^{2ig(-w)}\frac{\left[a n_{ag}^2/n_a^2 + i k_0 g(0)\right]}{\left[s n_{gs}^2/n_s^2 + i k_0 g(-w)\right]} = e^{i4\varphi^{TM}}\frac{\left[a n_{ag}^2/n_a^2 - i k_0 g(0)\right]}{\left[s n_{gs}^2/n_s^2 - i k_0 g(-w)\right]} \qquad \text{C-(205)}$$

or

$$e^{i4\varphi^{TM}} = e^{2ig(-w)}\,\frac{a n_{ag}^2/n_a^2 + i k_0 g(0)}{a n_{ag}^2/n_a^2 - i k_0 g(0)}\,\frac{s n_{gs}^2/n_s^2 - i k_0 g(-w)}{s n_{gs}^2/n_s^2 + i k_0 g(-w)} \qquad \text{C-(206)}$$

Using Eqs. B-(44), (194), and (195), then

$$e^{i4\varphi^{TM}} = e^{2ig(-w)}\, e^{-2i\varphi_{ag}^{TM}}\, e^{+2i\varphi_{gs}^{TM}} \qquad \text{C-(207)}$$

or

$$4\varphi^{TM} = 2g(-w) - 2\varphi_{ag}^{TM} + 2\varphi_{gs}^{TM} + 2m\pi \qquad \text{C-(208)}$$

Using Eq. (193), Eq. (208) implies that

$$\varphi^{TM} = -\varphi_{ag}^{TM} \qquad \text{C-(209)}$$

Turning now to A_1 and A_4, Eqs. (185) and (179) can be solved in terms of A_2. Using Eqs. (202) and (209)

$$A_1 = n_{ag}\, q^{-1/2}(0)\, A_2 \left(1 + e^{-2i\varphi_{ag}^{TM}}\right) \qquad \text{C-(210)}$$

$$= 2\, n_{ag}\, q^{-1/2}(0)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{ag}^{TM}\right) \qquad \text{C-(211)}$$

Also, from Eq. (179)

$$A_4 = e^{sW} n_{gs}\, q^{-1/2}(-W)\, A_2\, e^{-i\varphi_{ag}^{TM}} \left\{ e^{i\left[g(-W) + \varphi_{ag}^{TM}\right]} + e^{-i\left[g(-W) + \varphi_{ag}^{TM}\right]} \right\} \qquad \text{C-(212)}$$

or, from Eq. (193)

$$A_4 = e^{sW} n_{gs}\, q^{-1/2}(-W)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{gs}^{TM} + m\pi\right) \qquad \text{C-(213)}$$

Finally, from Eqs. (167), (202), and (209)

$$n_g\, q^{-1/2}\left[A_2 e^{ig} + A_3 e^{-ig}\right] = n_g\, q^{-1/2} A_2\, e^{-i\varphi_{ag}^{TM}} \left[e^{i(g + \varphi_{ag}^{TM})} + e^{-i(g + \varphi_{ag}^{TM})}\right] \qquad \text{C-(214)}$$

$$= 2 n_g\, q^{-1/2} A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left[g(x) + \varphi_{ag}^{TE}\right] \qquad \text{C-(215)}$$

It is now possible to write down the TM solutions for $\vec{E}$ and $\vec{H}$ by using Eqs. (142)-(147), together with Eqs. (139)-(141), (211), (212), and (215).

$$\vec{E}_{TM} = \hat{x}\, E_x + \hat{z}\, E_z \qquad \text{C-(216)}$$

$$\vec{H}_{TM} = \hat{y}\, H_y \qquad \text{C-(217)}$$

$$E_x(x,z,t) = E_{ox}(x)\, e^{i(\omega t - \beta z)} \qquad \text{C-(218)}$$

$$E_z(x,z,t) = E_{oz}(x)\, e^{i(\omega t - \beta z)} \qquad \text{C-(219)}$$

$$H_y(x,z,t) = H_{oy}(x)\, e^{i(\omega t - \beta z)} \qquad \text{C-(220)}$$

where

$$H_{oy}(x) = -2n_{ag}\, g^{-1/2}(0)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{ag}^{TM}\right) e^{-ax} \quad \text{for } x \geqq 0 \qquad \text{C-(221)}$$

$$H_{oy}(x) = -2n_g(x)\, g^{-1/2}(x)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left[g(x) + \varphi_{ag}^{TM}\right] \quad \text{for } -W \leqq x \leqq 0 \qquad \text{C-(222)}$$

$$H_{oy}(x) = -2n_{gs}\, g^{-1/2}(-W)\, A_2\, e^{-i\varphi_{ag}^{TM}} \cos\left(\varphi_{gs}^{TM} + m\pi\right) \quad \text{for } x \leqq -W \qquad \text{C-(223)}$$

From Eq. (142), $E_x = [i/k_o\, n^2(x)]\partial H_y/\partial z$ for all x; hence from Eqs. (220)-(223)

$$E_{ox}(x) = \left(\beta/k_o n_a^2\right) H_{oy}(x) \qquad \text{for } x \geqq 0 \qquad \text{C-(224)}$$

$$E_{ox}(x) = \left[\beta/k_o n_g^2(x)\right] H_{oy}(x) \qquad \text{for } -W \leqq x \leqq 0 \qquad \text{C-(225)}$$

$$E_{ox}(x) = \left(\beta/k_o n_s^2\right) H_{oy}(x) \qquad \text{for } x \leqq -W \qquad \text{C-(226)}$$

From Eq. (144), $E_z = [-i/k_o\, n^2(x)]\partial H_y/\partial x$ for all x, so from Eqs. (220)-(223) [recalling that $g'(x)/k_o = q(x)$],

$$E_{oz}(x) = i\left(a/k_o n_a^2\right) H_{oy}(x) \qquad \text{for } x \geqq 0 \qquad \text{C-(227)}$$

$$E_{oz}(x) = i\left[g(x)/n_g^2(x)\right] \tan\left[g(x) + \varphi_{ag}^{TM}\right] H_{oy}(x), \quad -W \leqq x \leqq 0 \qquad \text{C-(228)}$$

$$E_{oz}(x) = -i\left(s/k_o n_s^2\right) H_{oy}(x) \quad \text{for } x \leq -W \qquad \text{C-(229)}$$

where, in Eq. (228), terms of order $d[n_g(x)q^{-1/2}(x)]/dx$ are neglected according to Eq. (180) and the WKBJ approximation.

APPENDIX D

TOTAL INTERNAL REFRACTION IN INHOMOGENEOUS WAVEGUIDES

Introduction

In Appendix C it was shown that the WKBJ solutions for $\vec{E}$ and $\vec{H}$ are valid as long as Eqs. C-(71) and C-(175), for TE and TM modes respectively, are satisfied. When either Eq. C-(71) or C-(175) is not satisfied, it may be seen from Eqs. C-(132) through C-(135) and Eqs. C-(221) through C-(226) that at least two field components (E_{oy} and H_{ox} for the TE case, H_{oy} and E_{ox} for the TM case) become infinite as $q \to 0$ due to the factor $q^{-1/2}$. Since this is physically unacceptable, some other approach must be used to solve the wave equation near the level $q = 0$. In the field of radio ionospheric physics, Budden[11] and Wait[12] have developed a useful approach for describing the propagation of light near $q = 0$. This method will be adapted to optical waveguides.

First of all, it should be mentioned that $-W$ is now chosen to be the value of x for which $\Theta = \pi/2$ (and $q = 0$) for a given mode m and incident angle $\Theta(0)$, as shown in Figure IV-1. That is, at the level $x = -W$, $\Theta(-W) = \pi/2$, $q(-W) = 0$, and total internal refraction occurs.

The assumption is now made that, for a slowly varying index profile, q^2 is approximately linearly proportional to x near the level $q = 0$, i.e., near $x = -W$. This is nearly the same thing as saying that $n_g(x)$ is linear in this region, as may be seen from the following argument.

It is convenient to let

$$n_g(x) = n_o + b(x)\,\Delta n \qquad \text{D-(1)}$$

where

$$n_g(0) = n_o + \Delta n \qquad \text{D-(2)}$$

at the top of the guide, and

$$n_g(-D) = n_o \qquad \text{D-(3)}$$

where -D is the penetration depth of the index profile such that beyond x = -D there is no longer an index gradient. For instance, n_o might be the index of the film before ion implantation, and Δn the maximum index change (where $\Delta n \ll n_o$) which is generally a measurable quantity. The function b(x) thus gives the index profile, where

$$b(0) = 1 \qquad \text{D-(4)}$$

$$b(-D) = 0 \qquad \text{D-(5)}$$

Expanding $n_g(x)$ in a Taylor series expansion around x = -W, produces:

$$n_g(x) = n_g(-W) + (x+W)\,n_g'(-W) + \cdots \qquad \text{D-(6)}$$

$$= n_o + \Delta n\left[b(-W) + b'(-W)(x+W)\right] \qquad \text{D-(7)}$$

which is linear in x. For small Δn

$$n_g^2(x) \cong n_o^2 + 2n_o \Delta n \left[b(-W) + b'(-W)(x+W) \right] \qquad \text{D-(8)}$$

At the level q = 0 (and x = -W), θ has the value $\pi/2$, so

$$\eta^2 = n_g^2(-W) \sin^2\theta(-W) \qquad \text{D-(9)}$$

$$= n_g^2(-W) \qquad \text{D-(10)}$$

$$= n_o^2 + 2n_o \Delta n \, b(-W) \qquad \text{D-(11)}$$

hence [from Eq. C-(97)]

$$q^2(x) = \left[n_g^2(x) - \eta^2 \right] = 2n_o \Delta n \, b'(-W)(x+W) \qquad \text{D-(12)}$$

which is linear in x. Eq. (12) is equivalent to a Taylor expansion of $q^2(x)$ in the region near x = -W.

TE Modes

Using the form of q^2 given by Eq. (12) and the substitution[11]

$$\varsigma \equiv -\left[\frac{k_o}{(q^2)'} \right]^{2/3} q^2 \qquad \text{D-(13)}$$

where

$$(q^2)' = 2n_o \Delta n \, b'(-W) \qquad \text{D-(14)}$$

$$(q^2)'' = 0 \qquad \text{D-(15)}$$

then

$$\varsigma = -\left[\frac{k_o}{2 n_o \Delta n\, b'(-W)}\right]^{2/3} 2 n_o \Delta n\, b'(-W)(x+W) \qquad \text{D-(16)}$$

$$= -\left[2 k_o^2 n_o \Delta n\, b'(-W)\right]^{1/3} (x+W) \qquad \text{D-(17)}$$

and

$$\varsigma' = \frac{d\varsigma}{dx} = -\left[2 k_o^2 n_o \Delta n\, b'(-W)\right]^{1/3} \qquad \text{D-(18)}$$

$$\varsigma'' = \frac{d^2\varsigma}{dx^2} = 0 \qquad \text{D-(19)}$$

The wave equation A-(92) can be written as

$$\frac{\partial^2 F_x}{\partial x^2} + k_o^2 q^2(x) F_x = 0 \qquad \text{D-(20)}$$

since all z dependence of F_x is of the form $e^{-i\beta z}$, and $(-\beta^2 + k_o^2 n_g^2) = q^2$.

By the chain rule

$$\frac{\partial^2 F_x}{\partial x^2} = \frac{dF_x}{d\varsigma}\varsigma'' + \frac{d^2 F_x}{d\varsigma^2}(\varsigma')^2 \qquad \text{D-(21)}$$

$$= \frac{d^2 F_x}{d\varsigma^2}(\varsigma')^2 \qquad \text{D-(22)}$$

Hence, Eq. (20) becomes

$$\frac{d^2 F_x}{d\varsigma^2}(\varsigma')^2 + k_o^2 q^2 F_x = 0 \qquad \text{D-(23)}$$

which by Eqs. (18) and (12) is

$$\frac{d^2 F_x}{d\zeta^2}\left[2k_o^2 n_o \Delta n\, b'(-W)\right]^{2/3} + \left[2k_o^2 n_o \Delta n\, b'(-W)(x+W)\right]F_x = 0 \qquad \text{D-(24)}$$

or

$$\frac{d^2 F_x}{d\zeta^2} + \left[2k_o^2 n_o \Delta n\, b'(-W)\right]^{1/3}(x+W)\, F_x = 0 \qquad \text{D-(25)}$$

Using Eq. (17), Eq. (25)becomes the so-called Stokes differential equation:[11]

$$\frac{d^2 F_x}{d\zeta^2} - \zeta F_x = 0 \qquad \text{D-(26)}$$

Its solution as discussed by Budden[11] is the following:

$$F_x = K_{TE}\, e^{i(\omega t - \beta z)}\, Ai(\zeta) \qquad \text{D-(27)}$$

where K_{TE} is a constant and where Ai(ζ) is the Airy integral function defined as follows:

$$Ai(\zeta) = \frac{1}{\pi}\int_0^\infty \cos\left(\zeta u + \tfrac{1}{3}u^3\right) du \qquad \text{D-(28)}$$

Tables of Ai(ζ) are given by Miller.[15] A plot of Ai(ζ) versus ζ is given in Figure IV-4. It may be seen that beyond the region ζ = 0 (q = 0, x = -W) the solution Ai(ζ) is an exponentially damped wave--an evanescent wave.

There is another solution to Eq. (26) since it is a second-order differential equation. It is another Airy integral function Bi(ζ). However, beyond the level $x = -W$ at $q = 0$, the function Bi(ζ) becomes indefinitely large[11] as ζ increases, i.e., as $x \to -\infty$. Since it is physically unrealistic for F_x (and thus $\vec{E}$) to get larger and larger at distances farther and farther away from the waveguide, then the solution to Eq. (26) cannot contain a multiple of Bi(ζ).

For ζ sufficiently large, Ai(ζ) can be given by its asymptotic approximations[11]

$$Ai(\zeta) \cong \frac{1}{2}\frac{1}{\sqrt{\pi}}\frac{1}{\sqrt[4]{\zeta}}\left[e^{-\frac{2}{3}\zeta^{3/2}} + i\, e^{\frac{2}{3}\zeta^{3/2}}\right] \qquad \text{for } \tfrac{2}{3}\pi \leq \arg\zeta \leq \tfrac{4}{3}\pi \qquad \text{D-(29)}$$

and

$$Ai(\zeta) \cong \frac{1}{2}\frac{1}{\sqrt{\pi}}\frac{1}{\sqrt[4]{\zeta}}\, e^{-\frac{2}{3}\zeta^{3/2}} \qquad \text{for } -\tfrac{2}{3}\pi \leq \arg\zeta \leq \tfrac{2}{3}\pi \qquad \text{D-(30)}$$

where, for an exponent σ,

$$\zeta^{\sigma} = |\zeta|^{\sigma} e^{i\sigma \arg\zeta} \qquad \text{D-(31)}$$

From Eq. (17) it may be seen that

$$\arg\zeta = \pi \qquad \text{for } (x+W) > 0 \qquad \text{D-(32)}$$

$$\arg\zeta = 0 \qquad \text{for } (x+W) < 0 \qquad \text{D-(33)}$$

where Eqs. (29) and (32) correspond to propagating within the guide (where $q^2 > 0$, $-W < x < 0$), and where Eqs. (30) and (33) correspond to the evanescent wave region (where $q^2 < 0$, $x < -W$).

Referring to Figure IV-5 and following the arguments of Budden[11] and Wait,[12] it is assumed that q^2 is linear over the entire region be. There is a certain range cd where $|q|$ is so small that it violates the WKBJ approximation, and so the WKBJ solutions cannot be used. Outside this range the WKBJ solutions are good approximations. In the regions bc and de the asymptotic approximations of $Ai(\zeta)$ can be used, and they must be matched to the WKBJ solutions.

The WKBJ solutions are given from Eqs. C-(52) and C-(62):

$$F_{xg} = q^{-1/2}\,(1/i\beta)\,e^{i(\omega t-\beta z)}\left[F_2 e^{ik_0\int_0^x q\,dx} + F_3 e^{-ik_0\int_0^x q\,dx}\right] \qquad \text{D-(34)}$$

$$\text{for } -W \leq x \leq 0$$

$$F_{xs} = |q|^{-1/2}\,(1/i\beta)\,e^{i(\omega t-\beta z)}\,F_4\,e^{-k_0\int_{-W}^{x}|q|\,dx} \qquad \text{D-(35)}$$

$$\text{for } x \leq -W$$

Eq. (35) follows from Eq. (34) using the following argument. From Eq. C-(97) for $q(x)$, it may be seen that, because $n = n_g(x)\sin\Theta(x)$ can be set equal to $n_g(-W)$ [since $\Theta(-W) = \pi/2$], then $q(x)$ is imaginary for x beyond the point $x = -W$, i.e., $q = i|q|$. Since the $\vec{E}$ field cannot indefinitely increase as $x \to -\infty$, there can be only one term in Eq. (35) -- an exponentially decaying one.

In order to show the relationship between $Ai(\zeta)$ and the WKBJ solutions, Eq. (17) can be used. From Eqs. (31) and (32)

$$\zeta^{3/2} = -i\,|\zeta|^{3/2} \quad \text{for } \arg\zeta = \pi \qquad \text{D-(36)}$$

Thus

$$\frac{2}{3}\varsigma^{3/2} = -i\,\frac{2}{3}\left[2k_o^2 n_o \Delta n\, b'(-w)\right]^{1/2}(x+w)^{3/2} \qquad \text{D-(37)}$$

$$= -i\,\frac{2}{3}\left[2k_o^2 n_o \Delta n\, b'(-w)\right]^{1/2}\left[\frac{3}{2}\int_{-w}^{x}(x+w)^{1/2}dx\right] \qquad \text{D-(38)}$$

$$= -ik_o\int_{-w}^{x}\left[2n_o \Delta n\, b'(-w)\right]^{1/2}(x+w)^{1/2}dx \qquad \text{D-(39)}$$

which from Eq. (12) becomes

$$\frac{2}{3}\varsigma^{3/2} = -ik_o\int_{-w}^{x} q\,dx \quad \text{for } \arg\varsigma = \pi \qquad \text{D-(40)}$$

Eq. (40) can be expressed as follows:

$$\frac{2}{3}\varsigma^{3/2} = -ik_o\left[\int_{-w}^{0} q\,dx + \int_{0}^{x} q\,dx\right] \qquad \text{D-(41)}$$

From Eqs. (17), (31), and (32)

$$\sqrt[4]{\varsigma} = |\varsigma|^{1/4}\, e^{i\frac{1}{4}\arg\varsigma} \qquad \text{D-(42)}$$

$$= |\varsigma|^{1/4}\, e^{i\pi/4} \qquad \text{D-(43)}$$

$$= \left[2k_o^2 n_o \Delta n\, b'(-w)\right]^{1/12}(x+w)^{1/4}\, e^{i\pi/4} \qquad \text{D-(44)}$$

$$= \left[\frac{k_o}{2n_o \Delta n\, b'(-w)}\right]^{1/6}\left[2n_o \Delta n\, b'(-w)(x+w)\right]^{1/4} e^{i\pi/4} \qquad \text{D-(45)}$$

which from Eq. (12) becomes

$$\sqrt[4]{\zeta} = q^{+1/2}\left[\frac{k_o}{2n_o\Delta n\, b'(-W)}\right]^{1/6} e^{i\pi/4} \qquad \text{D-(46)}$$

Then for $\arg\zeta = \pi$, substitution of Eqs. (41) and (46) into Eq. (29) produces

$$Ai(\zeta) = q^{-1/2}\,\frac{1}{2}\,\frac{1}{\sqrt{\pi}}\left[\frac{2n_o\Delta n\, b'(-W)}{k_o}\right]^{1/6} e^{-i\pi/4}\left\{\exp\left[ik_o\int_{-W}^{0} q\,dx + ik_o\int_{0}^{x} q\,dx\right] + i\exp\left[-ik_o\int_{-W}^{0} q\,dx - ik_o\int_{0}^{x} q\,dx\right]\right\} \qquad \text{D-(47)}$$

The integral $-ik_o\int_{-W}^{0} q\,dx$ is a constant for a given W, and from Eq. C-(52), can be written as $ig(-W)$. Letting

$$K_1 \equiv \frac{1}{2}\,\frac{1}{\sqrt{\pi}}\left[\frac{2n_o\Delta n\, b'(-W)}{k_o}\right]^{1/6} \qquad \text{D-(48)}$$

then Eq. (47) can be written as

$$Ai(\zeta) = K_1\, q^{-1/2} e^{-i\pi/4} e^{-ig(-W)}\left[e^{ig(x)} + i\, e^{2ig(-W)} e^{-ig(x)}\right] \qquad \text{D-(49)}$$

for $\arg\zeta = \pi$

For $\arg\zeta = 0$,

$$\zeta^{1/4} = \left[\frac{k_o}{2n_o\Delta n\, b'(-W)}\right]^{1/6} q^{+1/2} \qquad \text{D-(50)}$$

$$\zeta^{3/2} = |\zeta|^{3/2} \qquad \text{D-(51)}$$

and, similar to Eq. (40)

$$\frac{2}{3}\zeta^{3/2} = k_o\int_{-W}^{x} |q|\,dx \qquad \text{for } \arg\zeta = 0 \qquad \text{D-(52)}$$

Substitution of Eqs. (50) and (52) into Eq. (30) produces

$$Ai(\zeta) = K_1 q^{-1/2} e^{-k_0 \int_{-W}^{x} |q| dx} \quad \text{for } \arg \zeta = 0 \qquad \text{D-(53)}$$

At the boundary of the linear q^2 region, the Airy integral solutions and the WKBJ solutions must match. Letting $x = -x_0$ be the location of the boundary between the regions bc and cd, then

$$\left[WKBJ\right]_{x=-x_0} = \left[Airy\right]_{x=-x_0} \qquad \text{D-(54)}$$

From Eqs. (27), (34), and (49)

$$\frac{1}{i\beta}\left[F_2 e^{ig(-x_0)} + F_3 e^{-ig(-x_0)}\right] =$$

$$K_{TE} K_1 e^{-i\pi/4} e^{-ig(-W)}\left[e^{ig(-x_0)} + i e^{2ig(-W)} e^{-ig(-x_0)}\right] \qquad \text{D-(55)}$$

From Eq. (55) using the orthogonality of the functions $e^{\pm ig}$, it can be concluded that

$$\frac{F_2}{i\beta} = K_{TE} K_1 e^{-i\pi/4} e^{-ig(-W)} \qquad \text{D-(56)}$$

$$\frac{F_3}{i\beta} = i K_{TE} K_1 e^{-i\pi/4} e^{+ig(-W)} \qquad \text{D-(57)}$$

or

$$K_{TE} = F_2 \left[e^{-i\pi/4} e^{ig(-W)} / \beta K_1\right] \qquad \text{D-(58)}$$

$$F_3 = F_2 \left[i e^{2ig(-W)}\right] \qquad \text{D-(59)}$$

It should be stated that equating the derivative of the WKBJ and the Airy integral solutions at the level $x = -x_0$ leads to the same results as Eqs. (58) and (59).

At the boundary between the regions cd and de, symmetrically located (with respect to x_0) on the other side of $x = -W$, the WKBJ and Airy solutions given by Eqs. (35) and (53) must also match. Letting $\Delta \equiv W - |x_0|$, and letting $x_1 = -W - \Delta$ be the location of this boundary, then

$$\frac{1}{\beta} F_4 e^{-k_0 \int_{-W}^{x_1} |g| dx} = K_{TE} K_1 e^{-k_0 \int_{-W}^{x_1} |g| dx} \qquad \text{D-(60)}$$

from which it may be concluded that

$$F_4 = \beta K_{TE} K_1 \qquad \text{D-(61)}$$

or, from Eq. (58)

$$F_4 = F_2 e^{-i\pi/4} e^{ig(-W)} \qquad \text{D-(62)}$$

At the $n_a - n_g$ (0) boundary, Eqs. C-(78) and C-(83) can be used for the boundary conditions. They yield Eq. C-(86) which can be written [with the aid of Eq. B-(44)] as

$$F_3 = F_2 e^{-2i\phi_{ag}^{TE}} \qquad \text{D-(63)}$$

where ϕ_{ag}^{TE} is defined in Eq. C-(92). Dividing Eq. (59) by Eq. (63) produces

$$1 = e^{-2im\pi} = i e^{2ig(-W)} e^{2i\phi_{ag}^{TE}} \qquad \text{D-(64)}$$

or

$$2g(-W) + 2\varphi_{ag}^{TE} + \frac{\pi}{2} = -2m\pi \qquad \text{D-(65)}$$

since $i = e^{i\pi/2}$. Using Eq. C-(53), Eq. (65) becomes

$$2k_o \int_{-W}^{0} n_g(x)\cos\theta(x)dx - 2\varphi_{ag}^{TE} - \frac{\pi}{2} = 2m\pi \qquad \text{D-(66)}$$

which is the waveguide equation for the TE case. Comparison of Eq. (66) with Eq. C-(96) shows that the only change is to replace the term $2\phi_{gs}^{TE}$ with $\pi/2$. Thus it may be seen that there is a phase advance of $\pi/2$ associated with the total internal refraction phenomenon. Further, from Eq. (66) it may be seen that the WKBJ solution predicts the correct phase change (except for the term $\pi/2$) experienced by the wave within the guide, even though the WKBJ solution itself breaks down in the region $q = 0$.

Using Eq. C-(97) and C-(98), Eq. (66) can be written as

$$2k_o \int_{-W}^{0} \left[n_g^2(x) - \eta^2\right]^{1/2} dx - 2\tan^{-1}\left\{\left[\frac{\eta^2 - n_a^2}{n_g^2(0) - \eta^2}\right]^{1/2}\right\} - \frac{\pi}{2} = 2m\pi \qquad \text{D-(67)}$$

Eq. (67) is trancendental in η, which is equivalent to saying that it is transcendental in $\theta(0)$, according to Eq. C-(101). It may or may not be transcendental in W too, depending on the functional form of $n_g(x)$ and the integral in Eq. (67).

Following arguments given in Appendix C, the cutoff condition is given when either

$$\eta \leq n_o \qquad \text{D-(68)}$$

or

$$\eta \leqq n_a \qquad \text{D-(69)}$$

For $n_o > n_a$ (which is usually the case), the cutoff condition is given by Eq. (68), i.e., when $\eta = n_o$. When $\eta < n_o$, the light can no longer be totally internally refracted, and the light escapes below the guide. For this cutoff condition, Eq. (67) becomes an integral equation for $(W_{min})_{TE}$:

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_o^2\right]^{1/2} dx = \frac{\lambda_o}{2\pi}\left\{ m\pi + \tan^{-1}\left[\left(\frac{n_o^2 - n_a^2}{2n_o\Delta n}\right)^{1/2}\right]\right\} \qquad \text{D-(70)}$$

where the $(\Delta n)^2$ term in the denominator of the argument of the arctangent has been neglected compared to $2n_o\Delta n$.

In the event that $n_o < n_a$ (a much less common situation) where $(n_a - n_o) < \Delta n$, then the cutoff condition is given when $\eta = n_a$ and the light leaks out into the superstrate. In this case $\phi_{ag}^{TE} = 0$. Thus Eq. (67) becomes

$$\int_{-(W_{min})_{TE}}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \lambda_o (4m+1)/8 \qquad \text{D-(71)}$$

Since $\sin\Theta(0)$ cannot exceed unity, then from Eq. C-(38) [or Eq. C-(101)], η can never exceed the value $n_g(0)$. In fact, η cannot equal $n_g(0)$ either. If $\sin\Theta(0) = 1$, then $\Theta(0) = \pi/2$, and from Snell's Law [Eq. C-(38)], $\Theta(x) = \pi/2$ for all x. Hence the integrand in Fq. (66) would be zero for all x, and in order to propagate a light wave, only an infinitely thick ($W \to \infty$) guide could satisfy Eq. (66), which is physically unrealistic. Thus the allowable values of η are

$$n_o + \Delta n > \eta > n_o \qquad \text{D-(72)}$$

in order that a wave may propagate in the guide (without cutoff). Using Eq. C-(101), this says that

$$1 > \sin\theta(0) > \frac{n_o}{n_o+\Delta n} \qquad \text{D-(73)}$$

which gives the allowable range of values of $\theta(0)$ for solving Eq. (67) when $n_o > n_a$. When $n_o < n_a$, then

$$n_o+\Delta n > \eta > n_a \qquad \text{D-(74)}$$

or

$$1 > \sin\theta(0) > \frac{n_a}{n_o+\Delta n} \qquad \text{D-(75)}$$

The effect of curvature of the index profile must be investigated in order to determine the validity of Eqs. (12) and (17) and the resultant conversion of the wave equation to the Stokes equation. If an additional term is kept in Eq. (6), then

$$n_g(x) = n_g(-w) + (x+w)n_g'(-w) + (x+w)^2 n_g''(-w)/2 + \cdots \qquad \text{D-(76)}$$

which, from Eq. (1) becomes

$$n_g(x) = n_o + \Delta n\left[b(-w) + b'(-w)(x+w) + b''(-w)(x+w)^2/2\right] \qquad \text{D-(77)}$$

$$n_g^2(x) \cong n_o^2 + 2n_o\Delta n\left[b(-w) + b'(-w)(x+w) + b''(-w)(x+w)^2/2\right] \qquad \text{D-(78)}$$

and

$$\eta^2 = n_g^2(-w) = n_o^2 + 2 n_o \Delta n\, b(-w) \qquad \text{D-(79)}$$

Hence

$$q^2 = n_g^2(x) - \eta^2 = 2 n_o \Delta n \left[b'(-w)(x+w) + b''(-w)(x+w)^2/2 \right] \qquad \text{D-(80)}$$

$$\equiv B_1 (x+w) + B_2 (x+w)^2 \qquad \text{D-(81)}$$

where

$$B_1 \equiv 2 n_o \Delta n\, b'(-w) \qquad \text{D-(82)}$$

$$B_2 \equiv n_o \Delta n\, b''(-w) \qquad \text{D-(83)}$$

Defining ζ according to Eq. (17), then

$$\zeta = -\left(k_o^2 B_1\right)^{1/3} (x+w) \qquad \text{D-(84)}$$

and the wave equation given by Eq. (20) becomes similar to Eq. (24) except for the additional quadratic term from Eq. (81):

$$\frac{d^2 F_x}{d\zeta^2} \left(k_o^2 B_1\right)^{2/3} + k_o^2 \left[B_1 (x+w) + B_2 (x+w)^2 \right] F_x = 0 \qquad \text{D-(85)}$$

or

$$\frac{d^2 F_x}{d\zeta^2} + \left[-\zeta + \zeta^2 \left(k_o B_1^2\right)^{-2/3} B_2 \right] F_x = 0 \qquad \text{D-(86)}$$

If $\mathrm{Ai}(\zeta)$ is to be a good approximation to the solution of Eq. (86), then

$$|\zeta| >> \left|\zeta^2 \left(k_o B_1^2\right)^{-2/3} B_2\right| \qquad \text{D-(87)}$$

and this must hold for values of ζ large enough for the asymptotic approximation of $\mathrm{Ai}(\zeta)$ (and thus the WKBJ solutions) to be valid. According to Budden,[11] this will happen when $|\zeta| \gtrsim 1$. So Eq. (87) requires that, for $|\zeta| = 1$, the linearity criterion is

$$|B_2| << \left|k_o B_1^2\right|^{2/3} \qquad \text{D-(88)}$$

From Eq. (1)

$$n_g'(x) = \Delta n\, b'(x) \qquad \text{D-(89)}$$

$$n_g''(x) = \Delta n\, b''(x) \qquad \text{D-(90)}$$

hence Eqs. (82), (83), and (88) require that

$$n_g''(-W) << \left(16\, n_o\, k_o^2\right)^{1/3} \left[n_g'(-W)\right]^{4/3} \qquad \text{D-(91)}$$

Further, from Eqs. (17) and (89), the condition that $|\zeta| = 1$ (at $x = -x_o$ and $x = -W - \Delta$, i.e., $x = -W \pm \Delta$) corresponds to

$$\Delta = |x + W| = \frac{1}{\left[2 k_o^2\, n_o\, n_g'(-W)\right]^{1/3}} \qquad \text{D-(92)}$$

It is now possible to write down the solutions for $\vec{E}_{TE}$ and $\vec{H}_{TE}$ using Eqs. B-(75) through B-(79) together with the results of this appendix and Appendix C. The solutions for $x \geq 0$ are given by Eqs. C-(132), C-(135), and C-(136). For the region $(-W + \Delta) \leq x \leq 0$, the solutions are the WKBJ solutions given by Eqs. C-(133), C-(135), and C-(137). For the region $(-W - \Delta) \leq x \leq (-W + \Delta)$, the solutions involve the Airy integral function. Using Eq. (27) together with Eqs. C-(5) through C-(10), then

$$E_{oy}(x) = i\beta K_{TE} \, Ai(\varsigma) \qquad \text{D-(93)}$$

$$H_{ox}(x) = -(\beta/k_o) E_{oy}(x,z,t) \qquad \text{D-(94)}$$

$$H_{oz}(x) = (i/k_o) \frac{\partial E_{oy}}{\partial \varsigma} \frac{\partial \varsigma}{\partial x} \qquad \text{D-(95)}$$

$$= (i/k_o)\left[\varsigma' \frac{\partial Ai(\varsigma)}{\partial \varsigma} \frac{1}{Ai(\varsigma)}\right] E_{oy}(x) \qquad \text{D-(96)}$$

where K_{TE} and ζ' are given by Eqs. (58) and (18) respectively, and where

$$\frac{d\,Ai(\varsigma)}{d\varsigma} = -\frac{1}{\pi}\int_0^\infty u \sin\left(\varsigma u + \tfrac{1}{3}u^3\right) du \qquad \text{D-(97)}$$

Values of $d[Ai(\zeta)]/d\zeta$ are tabulated by Miller.[15] Finally, for the region $x \leq (-W - \Delta)$, the solutions are the WKBJ solutions [obtained from Eq. (35)] and Eqs. C-(5) through C-(10):

$$E_{oy}(x) = -|q|^{-1/2} F_4 \, e^{-k_o \int_{-W}^{x} |q| dx} \qquad \text{D-(98)}$$

$$H_{ox}(x) = -(\beta/k_o)\, E_{oy}(x) \qquad \text{D-(99)}$$

$$H_{oz}(x) = -i\,|q|\, E_{oy}(x) \qquad \text{D-(100)}$$

where F_4 is given by Eq. (62), and where, in Eq. (100), terms order $d\left\{|q|^{-1/2}\right\}/dx$ are neglected according to Eqs. C-(81) and C-(82) and the WKBJ approximation.

TM Modes

The wave equation for TM modes is given by Eq. A-(93) and can be written as

$$\frac{\partial^2 A_x}{\partial x^2} - \frac{1}{n_g^2(x)}\,\frac{\partial A_x}{\partial x}\,\frac{\partial\left[n_g^2(x)\right]}{\partial x} + k_o^2 q^2(x) A_x = 0 \qquad \text{D-(101)}$$

Making the substitution[7,11]

$$U = A_x / n_g(x) \qquad \text{D-(102)}$$

then Eq. (101) becomes

$$\frac{\partial^2(U n_g)}{\partial x^2} - \frac{1}{n_g^2}\,\frac{\partial(U n_g)}{\partial x}\,\frac{\partial n_g^2}{\partial x} + k_o^2 q^2 U n_g = 0 \qquad \text{D-(103)}$$

or

$$U'' n_g + 2U' n_g' + U n_g'' - \frac{1}{n_g^2}\left(U' n_g + U n_g'\right) 2 n_g n_g' + k_o^2 q^2 U n_g = 0 \qquad \text{D-(104)}$$

$$U'' n_g + \left[n_g'' - 2\,\frac{(n_g')^2}{n_g} + k_o^2 q^2 n_g\right] U = 0 \qquad \text{D-(105)}$$

Thus, re-arranging terms, Eq. (105) becomes

$$\frac{\partial^2 U}{\partial x^2} + k_o^2 \left\{ q^2 - \frac{3}{4k_o^2 n_g^4} \left[\frac{\partial (n_g^2)}{\partial x} \right]^2 + \frac{1}{2k_o^2 n_g^2} \frac{\partial^2 (n_g^2)}{\partial x^2} \right\} U = 0 \qquad \text{D-(106)}$$

Defining an "effective refractive index"[11] as

$$N^2(x) \equiv n_g^2 - \frac{3}{4k_o^2 n_g^4} \left[\frac{\partial (n_g^2)}{\partial x} \right]^2 + \frac{1}{2k_o^2 n_g^2} \frac{\partial^2 (n_g^2)}{\partial x^2} \qquad \text{D-(107)}$$

and an "effective value of q"[11] as

$$Q^2(x) = N^2(x) - \eta^2 \qquad \text{D-(108)}$$

$$= q^2 - \frac{3}{4k_o^2 n_g^4} \left[\frac{\partial (n_g^2)}{\partial x} \right]^2 + \frac{1}{2k_o^2 n_g^2} \frac{\partial^2 (n_g^2)}{\partial x^2} \qquad \text{D-(109)}$$

then the wave equation becomes

$$\frac{\partial^2 U}{\partial x^2} + k_o^2 \, Q^2(x) \, U = 0 \qquad \text{D-(110)}$$

From Eqs. (107) - (109)

$$Q^2(x) - q^2(x) = N^2(x) - n_g^2(x) \qquad \text{D-(111)}$$

By making the assumption that $n_g(x)$ varies linearly near the level $x = -W$, as given by Eqs. (6) and (7), then

$$\frac{\partial (n_g^2)}{\partial x} = 2 n_o \Delta n \, b'(-W) \qquad \text{D-(112)}$$

$$\frac{\partial^2 (n_g^2)}{\partial x^2} = 0 \qquad \text{D-(113)}$$

$$\left[n_g^4(x)\right]^{-1} \cong \frac{1}{n_o^4}\left\{1 - \frac{4\Delta n}{n_o}\left[b(-w) + b'(-w)(x+w)\right]\right\} \quad \text{D-(114)}$$

Hence, from Eq. (109)

$$Q^2(x) - g^2(x) = -3\left[\frac{\Delta n\, b'(-w)}{k_o n_o}\right]^2\left\{1 - \frac{4\Delta n}{n_o}\left[b(-w) + b'(-w)(x+w)\right]\right\} \quad \text{D-(115)}$$

It may be seen that the difference between Q^2 and q^2 [and, using Eq. (111), between N^2 and n_g^2] is proportional to $[\Delta n\, b'(-W)/k_o]^2$, which is a very small quantity for optical waveguides. Hence, near the level $x = -W$, the wave equation Eq. (110) becomes approximately

$$\frac{\partial^2 U}{\partial x^2} + k_o^2 g^2(x) U = 0 \quad \text{D-(116)}$$

and all the techniques described for the TE case can be used to convert this form of the wave equation to the Stokes equation, and to then find the Airy integral solutions.

Neglecting the right-hand side of Eq. (115) is equivialent to the assumption that $Q^2(x)$ and $q^2(x)$ have a zero at nearly the same point. It is possible to calculate this slight difference between the TE and TM reflection points (i.e., the points of total internal refraction). It is convenient to define the TM reflection point as

$$W_{TM} \equiv W + \delta W \quad \text{D-(117)}$$

where W is the value of x at which q^2 is zero, and where δW is the difference. Using Eq. (12), Eq. (115) can then he written as

$$Q^2(x) = -Q_0 + (x + W_{TM} - \delta W)\, Q_1 \qquad \text{D-(118)}$$

where the constants Q_0 and Q_1 are defined as follows:

$$Q_0 \equiv 3\left\{\left[\frac{\Delta n\, b'(-W)}{k_0 n_0}\right]^2 \left[1 - \frac{4\Delta n\, b'(-W)}{n_0}\right]\right\} \qquad \text{D-(119)}$$

$$Q_1 \equiv \left\{2 n_0 \Delta n\, b'(-W) + \frac{12}{k_0^2}\left[\frac{\Delta n\, b'(-W)}{n_0}\right]^3\right\} \qquad \text{D-(120)}$$

In order that $Q(-W_{TM}) = 0$, then from Eq. (118) it is necessary that

$$\delta W = -\frac{Q_0}{Q_1} \qquad \text{D-(121)}$$

$$\cong -\frac{3}{2}\,\frac{\Delta n\, b'(-W)}{n_0^3 k_0^2} \qquad \text{D-(122)}$$

It may be seen that δW is indeed small for optical waveguides, and Eq. (116) is valid for TM modes to a very good approximation. Further, since $\delta W < 0$, then $W_{TM} < W$, i.e., a TM wave reflects at a value of $|x|$ which is smaller than in the TE case.

In Eq. (114), only the linear term was retained. If instead a quadratic term is also retained, then

$$\left[n_g^4(x)\right]^{-1} \cong \frac{1}{n_0^4}\left\{\left[1 - 4\Delta n\, b(-W)/n_0\right] - (x+W)\left[4\Delta n\, b'(-W)/n_0\right] + (x+W)^2\, 10\left[\Delta n\, b'(-W)/n_0\right]^2\right\} \qquad \text{D-(123)}$$

where it is assumed that terms containing the factor $b''(-W)$ are small enough according to the linearity criterion given by Eq. (88). Eq. (115) [or, equivalently, Eq. (109)] then becomes

$$Q^2(x) = -\left\{3\left[\frac{\Delta n\ b'(-w)}{k_0 n_0}\right]^2\left[1 - \frac{4\Delta n\, b(-w)}{n_0}\right]\right\}$$
$$+ (x+w)\left\{2n_0\,\Delta n\, b'(-w) + 12\,\frac{[\Delta n\, b'(-w)]^3}{n_0^3 k_0^2}\right\}$$
$$- (x+w)^2\left\{30\,\frac{[\Delta n\, b'(-w)]^4}{n_0^4 k_0^2}\right\} \qquad \text{D-(124)}$$

or

$$Q^2(x) = -Q_0 + (x + W_{TM} - \delta W)\,Q_1 - (x + W_{TM} - \delta W)^2\,Q_2 \qquad \text{D-(125)}$$

where the constant Q_2 is defined as

$$Q_2 \equiv \frac{30}{k_0^2}\left[\frac{\Delta n\ b'(-w)}{n_0}\right]^4 \qquad \text{D-(126)}$$

Collecting terms in Eq. (125) leads to

$$Q^2(x) = -\left[Q_0 + (\delta W)\,Q_1 + (\delta W)^2\,Q_2\right]$$
$$+ (x + W_{TM})\,(Q_1 + 2\delta W\,Q_2)$$
$$- (x + W_{TM})^2\,Q_2 \qquad \text{D-(127)}$$

Using Eq. (121) together with the fact that δW is small and $Q_2 << Q_1$, then Eq. (127) becomes to a good approximation

$$Q^2(x) = (x + W_{TM})\,Q_1 - (x + W_{TM})^2\,Q_2 \qquad \text{D-(128)}$$

which is of the same form as Eq. (81) for $q^2(x)$. Now, it is required that $Q^2(x)$ be linear near the region $x = -W$ [or, nearly the same, at the level $x = -W_{TM}$], just as it was required that q^2 be linear for the TE case. The quadratic term in Eq. (128) is subject to the linearity

criterion given by Eq. (88), i.e.,

$$|Q_2| << |k_o Q_1^2|^{2/3} \qquad \text{D-(129)}$$

Thus, from Eqs. (120) and (126)

$$|30\,[\Delta n\, b'(-w)/n_o]^4/k_o^2| << |k_o\,[2n_o \Delta n\, b'(-w)]^2|^{2/3} \qquad \text{D-(130)}$$

where the second term of Q_1 in Eq. (120) has been neglected in Eq. (130). Re-writing Eq. (130) leads to

$$n_o^2 >> \sqrt[8]{\frac{(15)^3}{2}}\ \Delta n\ b'(-w)/k_o \qquad \text{D-(131)}$$

Since the minimum value of η is n_o before cutoff, then Eq. (131) requires that

$$\eta^2 >> \sqrt[8]{\frac{(15)^3}{2}}\ \Delta n\ b'(-w)/k_o \qquad \text{D-(132)}$$

Budden[11] defines the quantity B as follows:

$$B \equiv \eta^2 \left\{ k_o/[n_g^2(x)] \right\}^{2/3}_{x=-W} \qquad \text{D-(133)}$$

[where optical waveguide notation has been used for the right-hand side of Eq. (133) instead of Budden's radio ionosphere notation]. Using Eq. (8), Eq. (133) becomes

$$B = \eta^2 \left[k_o/2n_o \Delta n\ b'(-w) \right]^{2/3} \qquad \text{D-(134)}$$

Again setting $n_o = \eta$, then Eq. (134) becomes

$$\eta = B^{3/4} \left[2 \Delta n \, b'(-W) / k_o\right]^{1/2} \qquad \text{D-(135)}$$

Using Eq. (135), Eq. (132) becomes

$$B >> \sqrt[4]{15/8} \qquad \text{D-(136)}$$

or

$$B >> 1.17 \qquad \text{D-(137)}$$

For optical waveguides, Eq. (134) almost always satisfies Eq. (137). That is, if the linearity criterion given by Eq. (88) is satisfied for TF modes, then it is satisfied for TM modes also, to an extremely good approximation. For instance, if the index profile is exponential such that

$$n_g(x) = n_o + \Delta n \left[1 - e^{d(x+D)}\right] \Big/ \left(1 - e^{dD}\right) \qquad \text{D-(138)}$$

then Eq. (138) satisfies Eqs. (1) - (3). If the following values are chosen: $\Delta n = 0.1$, $D = 1\mu m$ (which is a fairly steep profile), $\lambda_o = 0.63\mu m$, $d = 2/D = 2\mu m^{-1}$ (so that the e^{-1} point of the index profile occurs when $x = -D/2 = -0.5\mu m$); also, letting $\eta = n_o = 1.5$ (i.e., the wave is propagating nearly at cutoff), then $b'(-W) = b'(-D) = -d/(1 - e^{dD}) = 0.313$; for these values, then, from Eq. (134), $B = 51$, which certainly satisfies Eq. (137). For waveguide index profiles with $\Delta n = 0.01$ over a distance $D = 10\mu m$, B has the value of 1089.

Returning now to Eq. (125), the solution U has the form given by Eq. (27). Hence, from Eq. (102)

$$A_x = n_g \, K_{TM} \, e^{i(\omega t - \beta z)} \, Ai(S) \qquad \text{D-(139)}$$

This has to be matched to the WKBJ solutions at the levels $x = -W \pm \Delta$, as in the TE case. [Eq. (92) defines the value of Δ.] The WKBJ solutions are given from Fqs. C-(52) and C-(167):

$$A_{xg} = n_g q^{-1/2} (1/i\beta) e^{i(\omega t - \beta z)} \left[A_2 e^{ik_0 \int_0^x q \, dx} + A_3 e^{-ik_0 \int_0^x q \, dx} \right] \qquad \text{D-(140)}$$

for $-W \leq x \leq 0$

$$A_{xs} = n_g |q|^{-1/2} (1/i\beta) e^{i(\omega t - \beta z)} A_4 \, e^{-k_0 \int_{-W}^{x} |q| \, dx} \qquad \text{D-(141)}$$

for $x \leq -W$

Eqs. (139) - (141) differ from Fqs. (27), (34), and (35) only by the term $n_g(x)$. Hence, upon matching the boundary conditions at $x = -W \pm \Delta$ [and using the relations given by Eqs. (58), (59), and (62)],

$$K_{TM} = (A_2/F_2) \, K_{TE} \qquad \text{D-(142)}$$

$$A_3 = (A_2/F_2) \, F_3 \qquad \text{D-(143)}$$

$$A_4 = (A_2/F_2) \, F_4 \qquad \text{D-(144)}$$

At the n_a - $n_g(0)$ boundary

$$A_3 = A_2 \, e^{-2i\psi_{ag}^{TM}} \qquad \text{D-(145)}$$

similar to Eq. (63). [Fq. C-(194) defines the quantity ϕ_{ag}^{TM}]. Thus the waveguide equation for the TM case differs only slightly from the waveguide equation for the TE case, i.e.,

$$2k_o \int_{-W}^{0} n_g(x) \cos\theta(x) - 2\phi_{ag}^{TM} - \frac{\pi}{2} = 2m\pi \qquad \text{D-(146)}$$

similar to Eq. (66). Using Eqs. C-(97) and C-(194), Fq. (146) can be written as

$$2k_o \int_{-W}^{0} \left[n_g^2(x) - \eta^2\right]^{1/2} dx - 2\tan^{-1}\left\{\frac{n_g^2(0)}{n_a^2} \frac{\left(\eta^2 - n_a^2\right)^{1/2}}{\left[n_g^2(0) - \eta^2\right]^{1/2}}\right\} - \frac{\pi}{2} = 2m\pi \qquad \text{D-(147)}$$

similar to Eq. (67). The cutoff condition for $n_o > n_a$ is given by Eq. (68), in which case an integral equation for $(W_{min})_{TM}$ results:

$$\int_{-(W_{min})_{TM}}^{0} \left[n_g^2(x) - n_o^2\right]^{1/2} dx = \frac{\lambda_o}{2\pi}\left\{ m\pi + \tan^{-1}\left[\frac{n_o^2 + 2n_o\Delta n}{n_a^2}\left(\frac{n_o^2 - n_a^2}{2n_o\Delta n}\right)^{1/2}\right]\right\} \qquad \text{D-(148)}$$

similar to Fq. (70). In the event that $n_o < n_a$, then

$$\int_{-(W_{min})_{TM}}^{0} \left[n_g^2(x) - n_a^2\right]^{1/2} dx = \lambda_o (4m+1)/8 \qquad \text{D-(149)}$$

which is identical to Eq. (71). Again, the allowable values of η and $\theta(0)$ are given by Eqs. (72) - (75).

It is now possible to write down the TM solutions using Eqs. B-(149) through B-(153), Appendix C, and this appendix. For the region $x \geq 0$, the solutions are given by Eqs. C-(221), C-(224), and C-(227). For the region $(-W + \Delta) \leq x \leq 0$, the solutions are the WKBJ solutions given by Eqs. C-(222),

C-(225), and C-(228). For the region $(-W-\Delta) \leqq x \leqq (-W+\Delta)$, the solutions involve the Airy integral function [using Eqs. C-(142) through C-(147)]:

$$H_{oy}(x) = n_g \, i\beta K_{TE} \, Ai(\zeta) \qquad \text{D-(150)}$$

$$E_{ox}(x) = \left[\beta / k_o n_g^2(x)\right] \, H_{oy}(x) \qquad \text{D-(151)}$$

$$E_{oz}(x) = -\left[1/k_o n_g^2(x)\right]\left[\frac{n_g'}{n_g} + \zeta' \frac{\partial Ai(\zeta)}{\partial \zeta} \frac{1}{Ai(\zeta)}\right] H_{oy}(x) \qquad \text{D-(152)}$$

For the region $x \leqq (-W-\Delta)$, the solutions are the WKBJ solutions obtained from Eq. (141):

$$H_{oy}(x) = -n_g |q|^{-1/2} A_4 \, e^{-k_o \int_{-W}^{x} |q| \, dx} \qquad \text{D-(153)}$$

$$E_{ox}(x) = \left[\beta / k_o n_g^2(x)\right] \, H_{oy}(x) \qquad \text{D-(154)}$$

$$E_{oz}(x) = i\left[q(x)/n_g^2(x)\right] H_{oy}(x) \qquad \text{D-(155)}$$

where in Eq. (155), terms of order $d\left\{n_g |q|^{-1/2}\right\}/d_x$ are neglected according to Eqs. C-(181) and C-(182) and the WKBJ approximation.

APPENDIX E

THE PRINCIPAL AXES AND INDEXES OF CUBIC ZINC SULFIDE

The problem is to find the angle Ω through which the $\hat{y}$ and $\hat{z}$ axes must be rotated to the new axes $\hat{y}'$ and $\hat{z}'$, since, from Eqs. VI-(5) and VI-(6), $\hat{x}' = \hat{x}$ and $n_x' = n_o$. Referring to Figure E-1, $\hat{y}$ and $\hat{z}$ can be written as

$$\hat{y} = \hat{y}' \cos\Omega - \hat{z}' \sin\Omega \qquad \text{E-(1)}$$

$$\hat{z} = \hat{z}' \cos\Omega - \hat{y}' \sin\Omega \qquad \text{E-(2)}$$

Substituting Fqs. (1) and (2) into Eq. VI-(3) results in

$$(n_x')^{-2}\Big[x'^2 + (y'^2\cos^2\Omega + z'^2\sin^2\Omega - 2y'z'\sin\Omega\cos\Omega) + (z'^2\cos^2\Omega + y'^2\sin^2\Omega + 2y'z'\sin\Omega\cos\Omega) + 2r_{41}e_x(y'z'\cos^2\Omega + y'^2\cos\Omega\sin\Omega - z'^2\cos\Omega\sin\Omega - y'z'\sin^2\Omega = 1$$

E-(3)

Subtracting Eq. VI-(4) from Eq. (3), and using the fact that $\cos^2\Omega + \sin^2\Omega = 1$,

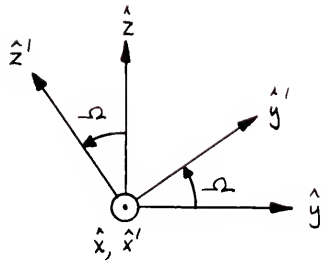


Figure E-1. Angular relationships between the $(\hat{x}, \hat{y}, \hat{z})$ set of axes and the $(\hat{x}', \hat{y}', \hat{z}')$ set.

$$+y'^2\left(\frac{1}{n_o^2}-\frac{1}{n_{y'}^2}+2r_{41}e_x\cos\Omega\sin\Omega\right)$$

$$+z'^2\left(\frac{1}{n_o^2}-\frac{1}{n_{z'}^2}-2r_{41}e_x\cos\Omega\sin\Omega\right)$$

$$+y'z'2r_{41}e_x\left(\cos^2\Omega-\sin^2\Omega\right)=0 \qquad \text{E-(4)}$$

The angle Ω is that angle which causes the y' z' terms to vanish, i.e.,

$$y'z'2r_{41}e_x\left(\cos^2\Omega-\sin^2\Omega\right)=0 \qquad \text{E-(5)}$$

For arbitrary e_x, this requires that $\cos^2\Omega = \sin^2\Omega$; thus

$$\Omega = \pi/4 + j\,\pi/2\,, \quad j=0,1,2,\cdots \qquad \text{E-(6)}$$

It is convenient to use j = 0 in Eq.(6) with no loss of generality. The new axes $\hat{x}'$, $\hat{y}'$, and $\hat{z}'$ -- the principal axes[26] -- are shown in Figure VI-1.

Because the $\hat{y}'$ and $\hat{z}'$ axes are orthogonal, the y'^2 and z'^2 terms in Eq. (4) must each be zero. Using $\Omega = \pi/4$, this requires that

$$y'^2\left(\frac{1}{n_o^2}-\frac{1}{n_{y'}^2}+r_{41}e_x\right)=0 \qquad \text{E-(7)}$$

$$z'^2\left(\frac{1}{n_o^2}-\frac{1}{n_{z'}^2}-r_{41}e_x\right)=0 \qquad \text{E-(8)}$$

Solving Eqs. (7) and (8) then

$$\frac{1}{n_y'^2} = \frac{1}{n_o^2} + r_{41} e_x \qquad \text{E-(9)}$$

$$\frac{1}{n_z'^2} = \frac{1}{n_o^2} - r_{41} e_x \qquad \text{E-(10)}$$

Letting

$$n_y' \equiv n_o + \Delta n_y' \qquad \text{E-(11)}$$

$$n_z' \equiv n_o + \Delta n_z' \qquad \text{E-(12)}$$

then Eqs. (9) and (10) can be written as

$$\frac{1}{(n_o + \Delta n_y)^2} = \frac{1}{n_o^2} + r_{41} e_x \qquad \text{E-(13)}$$

$$\frac{1}{(n_o + \Delta n_z)^2} = \frac{1}{n_o^2} - r_{41} e_x \qquad \text{E-(14)}$$

In general, the index changes Δn_y and Δn_z are very much less than n_o such that $\Delta n_y'/n_o \ll 1$ and $\Delta n_z'/n_o \ll 1$. So Eqs. (13) and (14) can be written as

$$\frac{1}{n_o^2}\left(1 - \frac{2\Delta n_y'}{n_o}\right) = \frac{1}{n_o^2}\left(1 + n_o^2 r_{41} e_x\right) \qquad \text{E-(15)}$$

$$\frac{1}{n_o^2}\left(1 - \frac{2\Delta n_z'}{n_o}\right) = \frac{1}{n_o^2}\left(1 - n_o^2 r_{41} e_x\right) \qquad \text{E-(16)}$$

Thus

$$\Delta n_y' = -n_o^3 r_{41} e_x/2 \qquad \text{E-(17)}$$

$$\Delta n_z' = + n_o^3 r_{41} e_x/2 \qquad \text{E-(18)}$$

Substituting Eqs. (17) and (18) into Eq. (11) and (12) [and repeating Eq. VI-(6)], the principal indexes are

$$n_x' = n_o \qquad \text{E-(19)}$$

$$n_y' = n_o - n_o^3 r_{41} e_x/2 \qquad \text{E-(20)}$$

$$n_z' = n_o + n_o^3 r_{41} e_x/2 \qquad \text{E-(21)}$$

LIST OF REFERENCES

1. R. F. Harington, Time-Harmonic Electromagnetic Fields (New York, McGraw-Hill, 1961), pp. 98-99, 129-131, 163-168.
2. H. Kirchoff, Arch. Elektron. & Uebetragungtsech. 26, 537 (1972).
3. P. K. Tien and R. Ulrich, J. Opt. Soc. Am. 60, 1325 (1970).
4. M. Born and E. Wolf, Principles of Optics (New York, Pergamon, 1970), 4th ed., p. 49.
5. Ibid., p. 110.
6. J. A. Stratton, Electromagnetic Theory (New York, McGraw-Hill, 1941), pp. 35-37.
7. R. Jacobsson, "Light Reflection from Films of Continuously Varying Refractive Index," Chapter V, Progress in Optics (Amsterdam, North-Holland Publishing Co., 1966), vol. V, ed. by E. Wolf, pp.247-286.
8. J. R. Wait, Electromagnetic Waves in Stratified Media (New York, Pergamon, 1970), pp. 86-95.
9. S. Lipson and H. Lipson, Optical Physics (London, Cambridge University Press, 1969), pp. 88-93.
10. K. G. Budden, Radio Waves in the Ionosphere (London, Cambridge University Press, 1961), pp. 140-143.
11. Ibid., pp. 283-352.
12. Wait, pp. 95-100, 338.
13. A Sommerfield, Optics (New York, Academic, 1964), pp. 322-324.
14. D. Marcuse, IEEE J. Quantum Electron QE-9, 1000 (1973).
15. J. C. P. Miller, "The Airy Integral," Brit. Assoc. Math. Tables (London, Cambridge University Press, 1946), vol. B.

16. S. S. Ballard, J. S. Browder, and J. F. Ebersole, "Refractive Index of Special Crystals and Certain Glasses," Section 6b, American Institute of Physics Handbook (New York, McGraw-Hill, 1972), 3rd ed., ed. by D. E. Gray.

17. M. E. Behrndt and S. C. Moreno, J. Vacuum Sci. Technol. 8, 494 (1971).

18. P. Vohl, W. R. Buchan, and J. E. Genthe, J. Electrochem. Soc. 118, 1842 (1971).

19. H. R. Philipp and H. Ehrenreich, Phys. Rev. 129, 1550 (1963).

20. F. A. Jenkins and H. E. White, Fundamentals of Optics (New York, McGraw-Hill, 1957), 3rd ed., pp. 475, 522.

21. Born and Wolf, pp. 59-63, 627-630.

22. Data Sheet, "Substrates for Epitaxy," Central Research Laboratories, Itek Corporation, 1970.

23. W. R. Oliver, Phil. Mag. 21, 1229 (1970).

24. M. Hacskaylo and C. Feldman, J. Appl. Phys. 33, 3042 (1962).

25. R. Ulrich and W. Prettl, Appl. Phys. 1, 55 (1973).

26. S. Namba, J. Opt. Soc. Am. 51, 76 (1961).

27. D. S. Oliver and W. R. Buchan, IEEE Trans. Electron Devices ED-18, 769 (1971).

28. R. J. Friauf, K. F. Young, W. J. Merz, "Properties of Ionic Crystals," Section 9f, AIP Handbook, op. cit.

29. W. E. Martin, J. Appl. Phys. 44, 3703 (1973).

30. D. B. Hall and C. Yeh, J. Appl. Phys. 44, 2271 (1973).

31. D. C. Doughty and J. M. Ley, Electron. Lett. 6, 141 (1970).

BIOGRAPHICAL SKETCH

John Franklin Ebersole was born on February 21, 1946 in Boston, Mass., the son of Mr. and Mrs. Laurence D. Ebersole. He graduated from Lunenburg (Mass.) High School in June 1964. He received a Bachelor of Arts degree with a major in Physics from Holy Cross College in May 1968, and a Master of Science degree (with thesis), major in Physics and minor in Astronomy, from the University of Florida in December 1971. For his Ph.D. degree, Mr. Ebersole has two minors: Astronomy and Computer & Information Sciences.

Mr. Ebersole has worked as a research assistant in laser interferometry at Holy Cross College, as an optical engineer in infrared and visible holography at the American Optical Company, and as a teaching assistant in optics in the Department of Physics of the University of Florida. Since March 1972, Mr. Ebersole has been employed as an optical engineer by the Optical Systems Division of Itek Corporation; he is currently involved in advanced optical technology research.

Mr. Ebersole is a member of the Optical Society of America, the Society of Photo-Optical Instrumentation Engineers, Sigma Xi, and Sigma Pi Sigma; he is a past president of the Student Physics Society chapters at Holy Cross College and at the University of Florida. He was a joint editor of two of the optical materials sections in the third edition of the American Institute of Physics Handbook, was an editor of the Journal of the Holy Cross Physics Society, and is a co-author of various articles in Applied Optics and of three patents filed or soon to be filed by Itek Corporation.

Mr. Ebersole is married to the former Ingrid Elizabeth Arntsen. They have two children, Kirsten Marie and John Franklin, Jr.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stanley S. Ballard, Chairman
Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Kwan-Yu Chen
Associate Professor of Astronomy

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James B. Conklin, Jr.
Associate Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ralph C. Isler
Associate Professor of Physics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Frank D. Vickers
Associate Professor of Computer and Information Sciences

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and qualtiy, as a dissertation for the degree of Doctor of Philosophy.

Gray Ward

Gray D. Ward
Assistant Professor of Physics

This dissertation was submitted to the Graduate Faculty of the Department of Physics and Astronomy in the College of Arts and Sciences and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

March, 1974

Dean, Graduate School



UNIVERSITY OF FLORIDA
3 1262 08666 207 8